CALCULUS ILLUSTRATED

VOLUME 4: CALCULUS IN HIGHER DIMENSIONS

PETER SAVELIEV

To the student

Mathematics is a science. Just as the rest of the scientists, mathematicians are trying to understand how the Universe operates and discover its laws. When successful, they write these laws as short statements called "theorems". In order to present these laws conclusively and precisely, a dictionary of the new concepts is also developed; its entries are called "definitions". These two make up the most important part of any mathematics book.

This is how definitions, theorems, and some other items are used as building blocks of the scientific theory we present in this text.

Every new concept is introduced with utmost specificity.

Definition 0.0.1: square root

Suppose a is a positive number. Then the *square root* of a is a positive number x, such that $x^2 = a$.

The term being introduced is given in *italics*. The definitions are then constantly referred to throughout the text.

New symbolism may also be introduced.

Square root

$$\sqrt{a}$$

Consequently, the notation is freely used throughout the text.

We may consider a specific instance of a new concept either before or after it is explicitly defined.

Example 0.0.2: length of diagonal

What is the length of the diagonal of a 1×1 square? The square is made of two right triangles and the diagonal is their shared hypotenuse. Let's call it a. Then, by the *Pythagorean Theorem*, the square of a is $1^2 + 1^2 = 2$. Consequently, we have:

$$a^2 = 2\,.$$

We immediately see the need for the square root! The length is, therefore, $a = \sqrt{2}$.

You can skip some of the examples without violating the flow of ideas, at your own risk.

All new material is followed by a few little tasks, or questions, like this.

Exercise 0.0.3

Find the height of an equilateral triangle the length of the side of which is 1.

The exercises are to be attempted (or at least considered) immediately.

Most of the in-text exercises are not elaborate. They aren't, however, entirely routine as they require understanding of, at least, the concepts that have just been introduced. Additional exercise *sets* are placed in the appendix as well as at the book's website: calculus123.com. Do not start your study with the exercises! Keep in mind that the exercises are meant to test – indirectly and imperfectly – how well the *concepts* have been learned.

There are sometimes words of caution about common mistakes made by the students.

Warning!

In spite of the fact that $(-1)^2 = 1$, there is only one square root of 1, $\sqrt{1} = 1$.

The most important facts about the new concepts are put forward in the following manner.

Theorem 0.0.4: Product of Roots

For any two positive numbers a and b, we have the following identity:

$$\boxed{\sqrt{a} \cdot \sqrt{b} = \sqrt{a \cdot b}}$$

The theorems are constantly referred to throughout the text.

As you can see, theorems may contain formulas; a theorem supplies limitations on the applicability of the formula it contains. Furthermore, every formula is a part of a theorem, and using the former without knowing the latter is perilous.

There is no need to memorize definitions or theorems (and formulas), initially. With enough time spent with the material, the main ones will eventually become familiar as they continue to reappear in the text. Watch for words “important”, “crucial”, etc. Those new concepts that do not reappear in this text are likely to be seen in the next mathematics book that you read. You need to, however, be aware of all of the definitions and theorems and be able to find the right one when necessary.

Often, but not always, a theorem is followed by a thorough argument as a justification.

Proof.

Suppose $A = \sqrt{a}$ and $B = \sqrt{b}$. Then, according to the definition, we have the following:

$$a = A^2 \quad \text{and} \quad b = B^2 \, .$$

Therefore, we have:

$$a \cdot b = A^2 \cdot B^2 = A \cdot A \cdot B \cdot B = (A \cdot B) \cdot (A \cdot B) = (AB)^2 \, .$$

Hence, $\sqrt{ab} = A \cdot B$, again according to the definition.

Some proofs can be skipped at first reading.

Its highly detailed exposition makes the book a good choice for *self-study*. If this is your case, these are my suggestions.

While reading the book, try to make sure that you understand new concepts and ideas. Keep in mind, however, that some are more important that others; they are marked accordingly. Come back (or jump forward) as needed. Contemplate. Find other sources if necessary. You should not turn to the exercise sets until you have become comfortable with the material.

What to do about exercises when solutions aren’t provided? First, use the examples. Many of them contain a problem – with a solution. Try to solve the problem – before or after reading the solution. You can also find exercises online or make up your own problems and solve them!

I strongly suggest that your solution should be thoroughly *written*. You should write in complete sentences, including all the algebra. For example, you should appreciate the difference between these two:

Wrong: $\boxed{\begin{array}{l} 1+1 \\ 2 \end{array}}$ Right: $\boxed{\begin{array}{l} 1+1 \\ = 2 \end{array}}$

The latter reads “one added to one is two”, while the former cannot be read. You should also justify all your steps and conclusions, including all the algebra. For example, you should appreciate the difference between these two:

Wrong: $\boxed{\begin{array}{l} 2x = 4 \\ x = 2 \end{array}}$ Right: $\boxed{\begin{array}{l} 2x = 4\,;\ \text{therefore,} \\ x = 2\,. \end{array}}$

The standards of thoroughness are provided by the examples in the book.

Next, your solution should be thoroughly *read*. This is the time for self-criticism: Look for errors and weak spots. It should be re-read and then rewritten. Once you are convinced that the solution is correct and the presentation is solid, you may show it to a knowledgeable person for a once-over.

Next, you may turn to modeling projects. Spreadsheets (Microsoft Excel or similar) are chosen to be used for graphing and modeling. One can achieve as good results with packages specifically designed for these purposes, but spreadsheets provide a tool with a wider scope of applications. Programming is another option.

Good luck!

August 8, 2020

To the teacher

The bulk of the material in the book comes from my lecture notes.

There is little emphasis on closed-form computations and algebraic manipulations. I do think that a person who has never integrated by hand (or differentiated, or applied the quadratic formula, etc.) cannot possibly understand integration (or differentiation, or quadratic functions, etc.). However, a large proportion of time and effort can and should be directed toward:

- understanding of the concepts and
- modeling in realistic settings.

The challenge of this approach is that it requires more abstraction rather than less.

Visualization is the main tool used to deal with this challenge. Illustrations are provided for every concept, big or small. The pictures that come out are sometimes very precise but sometimes serve as mere metaphors for the concepts they illustrate. The hope is that they will serve as visual "anchors" in addition to the words and formulas.

It is unlikely that a person who has never plotted the graph of a function by hand can understand graphs or functions. However, what if we want to plot more than just a few points in order to visualize curves, surfaces, vector fields, etc.? Spreadsheets were chosen over graphic calculators for visualization purposes because they represent the shortest step away from pen and paper. Indeed, the data is plotted in the simplest manner possible: one cell - one number - one point on the graph. For more advanced tasks such as modeling, spreadsheets were chosen over other software and programming options for their wide availability and, above all, their simplicity. Nine out of ten, the spreadsheet shown was initially created from scratch in front of the students who were later able to follow my footsteps and create their own.

About the tests. The book isn't designed to prepare the student for some preexisting exam; on the contrary, assignments should be based on what has been learned. The students' understanding of the concepts needs to be tested but, most of the time, this can be done only indirectly. Therefore, a certain share of routine, mechanical problems is inevitable. Nonetheless, no topic deserves more attention just because it's likely to be on the test.

If at all possible, don't make the students memorize formulas.

In the order of topics, the main difference from a typical calculus textbook is that sequences come before everything else. The reasons are the following:

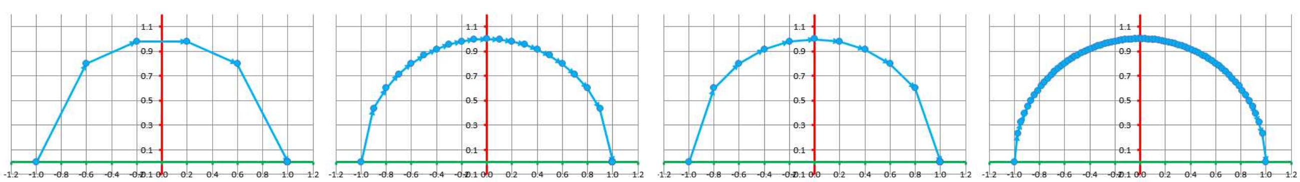
- Sequences are the simplest kind of functions.
- Limits of sequences are simpler than limits of general functions (including the ones at infinity).
- The sigma notation, the Riemann sums, and the Riemann integral make more sense to a student with a solid background in sequences.
- A quick transition from sequences to series often leads to confusion between the two.
- Sequences are needed for modeling, which should start as early as possible.

From the discrete to the continuous

It's no secret that a vast majority of calculus students will never use what they have learned. Poor career choices aside, a former calculus student is often unable to recognize the mathematics that is supposed to surround him. Why does this happen?

Calculus is the science of change. From the very beginning, its peculiar challenge has been to study and measure *continuous* change: curves and motion along curves. These curves and this motion are represented by *formulas*. Skillful manipulation of those formulas is what solves calculus problems. For over 300 years, this approach has been extremely successful in sciences and engineering. The successes are well-known: projectile motion, planetary motion, flow of liquids, heat transfer, wave propagation, etc. Teaching calculus follows this approach: An overwhelming majority of what the student does is manipulation of formulas on a piece of paper. But this means that all the problems the student faces were (or could have been) solved in the 18th or 19th centuries!

This isn't good enough anymore. What has changed since then? The computers have appeared, of course, and computers don't manipulate formulas. They don't help with solving – in the traditional sense of the word – those problems from the past centuries. Instead of *continuous*, computers excel at handling *incremental* processes, and instead of formulas they are great at managing discrete (digital) data. To utilize these advantages, scientists "discretize" the results of calculus and create algorithms that manipulate the digital data. The solutions are approximate but the applicability is unlimited. Since the 20th century, this approach has been extremely successful in sciences and engineering: aerodynamics (airplane and car design), sound and image processing, space exploration, structure of the atom and the universe, etc. The approach is also circuitous: Every concept in calculus *starts* – often implicitly – as a discrete approximation of a continuous phenomenon!



Calculus is the science of change, *both* incremental and continuous. The former part – the so-called discrete calculus – may be seen as the study of incremental phenomena and the quantities *indivisible* by their very nature: people, animals, and other organisms, moments of time, locations of space, particles, some commodities, digital images and other man-made data, etc. With the help of the calculus machinery called "limits", we invariably choose to transition to the continuous part of calculus, especially when we face continuous phenomena and the quantities *infinitely divisible* either by their nature or by assumption: time, space, mass, temperature, money, some commodities, etc. Calculus produces definitive results and absolute accuracy – but only for problems amenable to its methods! In the classroom, the problems are simplified until they become manageable; otherwise, we circle back to the discrete methods in search of approximations.

Within a typical calculus course, the student simply never gets to complete the "circle"! Later on, the graduate is likely to think of calculus only when he sees formulas and rarely when he sees numerical data.

In this book, every concept of calculus is first introduced in its discrete, "pre-limit", incarnation – elsewhere typically hidden inside proofs – and then used for modeling and applications well before its continuous counterpart emerges. The properties of the former are discovered first and then the matching properties of the latter are found by making the increment smaller and smaller, at the *limit*:

$$\boxed{\begin{array}{c}\text{discrete}\\ \text{calculus}\end{array}} \xrightarrow{\Delta x\to 0} \boxed{\begin{array}{c}\text{continuous}\\ \text{calculus}\end{array}}$$

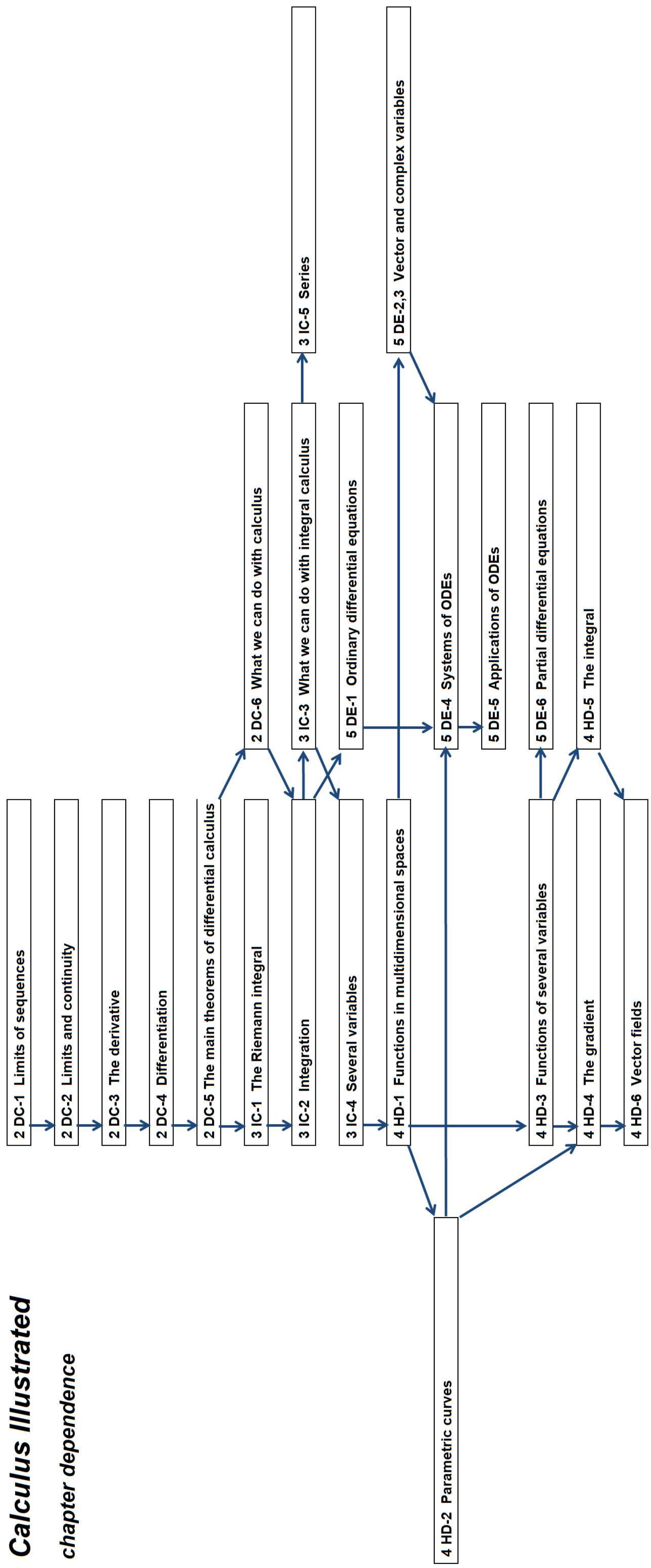
The volume and chapter references for *Calculus Illustrated*

This book is a part of the series *Calculus Illustrated*. The series covers the standard material of the undergraduate calculus with a substantial review of precalculus and a preview of elementary ordinary and partial differential equations. Below is the list of the books of the series, their chapters, and the way the present book (parenthetically) references them.

	■ **Calculus Illustrated. Volume 1: Precalculus**
1 PC-1	Calculus of sequences
1 PC-2	Sets and functions
1 PC-3	Compositions of functions
1 PC-4	Classes of functions
1 PC-5	Algebra and geometry
	■ **Calculus Illustrated. Volume 2: Differential Calculus**
2 DC-1	Limits of sequences
2 DC-2	Limits and continuity
2 DC-3	The derivative
2 DC-4	Differentiation
2 DC-5	The main theorems of differential calculus
2 DC-6	What we can do with calculus
	■ **Calculus Illustrated. Volume 3: Integral Calculus**
3 IC-1	The Riemann integral
3 IC-2	Integration
3 IC-3	What we can so with integral calculus
3 IC-4	Several variables
3 IC-5	Series
	■ **Calculus Illustrated. Volume 4: Calculus in Higher Dimensions**
4 HD-1	Functions in multidimensional spaces
4 HD-2	Parametric curves
4 HD-3	Functions of several variables
4 HD-4	The gradient
4 HD-5	The integral
4 HD-6	Vector fields
	■ **Calculus Illustrated. Volume 5: Differential Equations**
5 DE-1	Ordinary differential equations
5 DE-2	Vector variables
5 DE-3	Vector and complex variables
5 DE-4	Systems of ODEs
5 DE-5	Applications of ODEs
5 DE-6	Partial differential equations

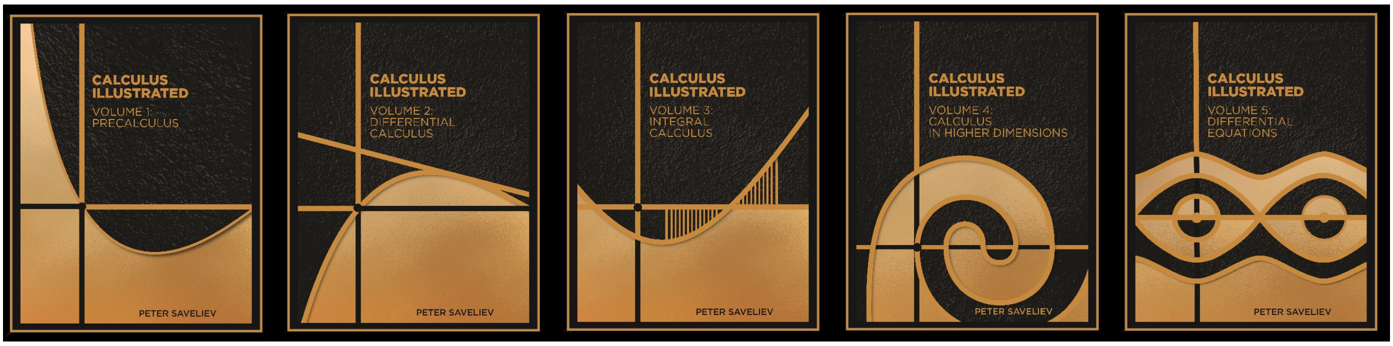
Each volume can be read independently.

A possible sequence of chapters is presented below. An arrow from A to B means that chapter B shouldn't be read before chapter A.



About the author

Peter Saveliev is a professor of mathematics at Marshall University, Huntington, West Virginia, USA. After a Ph.D. from the University of Illinois at Urbana-Champaign, he devoted the next 20 years to teaching mathematics. Peter is the author of a graduate textbook *Topology Illustrated* published in 2016. He has also been involved in research in algebraic topology and several other fields. His non-academic projects have been: digital image analysis, automated fingerprint identification, and image matching for missile navigation/guidance.



Contents

Chapter 1: Functions in multidimensional spaces

Contents

1.1. Multiple variables, multiple dimensions

Why do we need to study multidimensional spaces?

These are the main sources of spaces of *multiple dimensions*:

1. The physical space, dimension 3.
2. Multiple spaces of single dimension interconnected via functional relations: The graphs of these functions lie in higher-dimensional spaces.
3. Multiple quantities, homogeneous (such as stock and commodity prices) and non-homogeneous (such as other data): They are combined into points in abstract spaces.

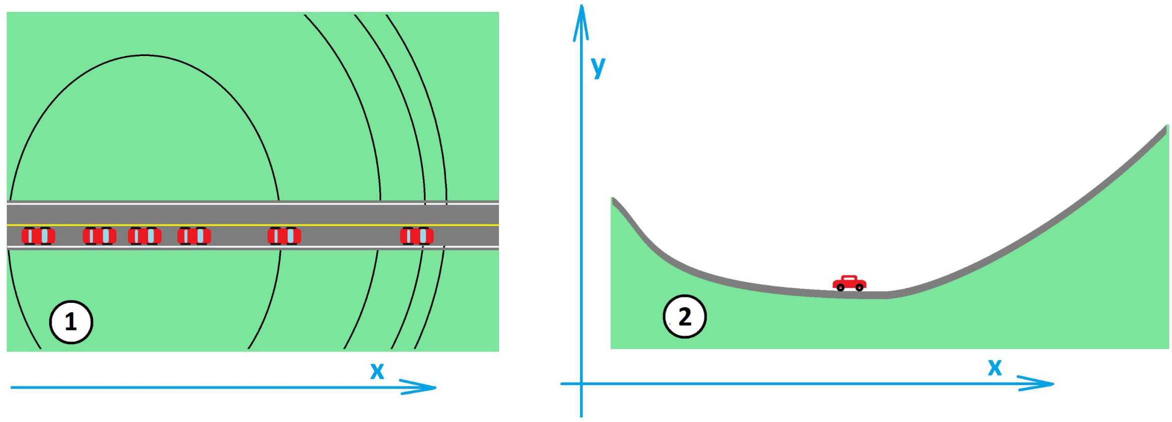
The 3-dimensional space represents a significant challenge in comparison to the plane. Furthermore, taking into account time will make it 4-dimensional.

Furthermore, planning a flight of a *plane* would require 3 spatial variables, but the number increases to 6 if we are to consider the orientation of the plane in the air: the roll, the pitch, and the yaw.

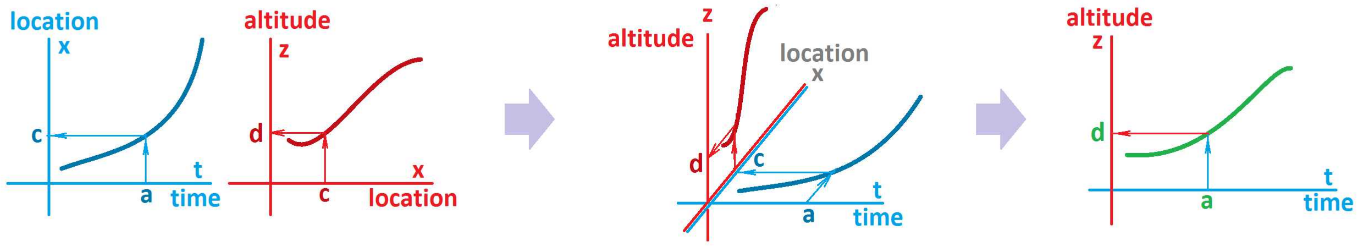
Next, let's notice that even when we deal with only numerical functions, the graph of such a function lies in the xy-plane, a space of dimension 2. What if there are two such functions?

Example 1.1.1: road trip

Let's imagine a car driven through a mountain terrain. Its location and its speed, as seen on the map, are known. The grade of the road is also known. How fast is the car climbing?



The first variable is time, t. We also have two *spatial* variables: the horizontal location x and the elevation (the vertical location) z. Then z depends on x, and x depends on t. Therefore, z depends on t via the *composition*:



Plotting both functions together requires a 3-dimensional space.

We can take specific functions:

- The horizontal location is a *linear* function of time, $x = 2t - 1$.
- The elevation is a *linear* function of horizontal location, $z = 3x + 7$.
- Then elevation is, too, a *linear* function of time, $z = 3(2t - 1) + 7$.

We can now answer the question directly:

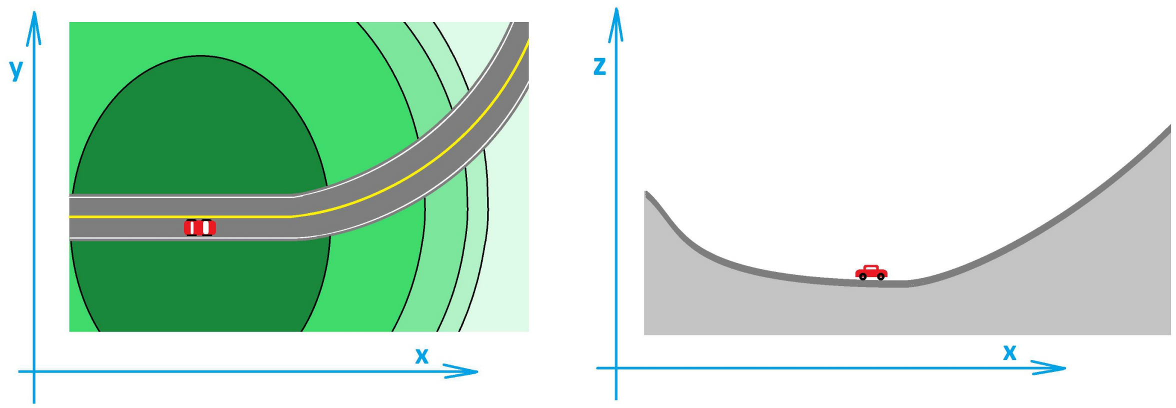
$$\frac{dz}{dt} = \frac{d}{dt}((2t)^2) = 8t\,,$$

or via the Chain Rule:

$$\frac{dz}{dt} = \frac{dz}{dx} \cdot \frac{dx}{dt} = \frac{d}{dx}(x^2) \cdot \frac{d}{dt}(2t) = 2x \cdot 2 = 2(2t) \cdot 2 = 8t\,.$$

Example 1.1.2: another road trip

Now, the road in this example is straight! More realistically, it should have turns and curves...



Fortunately, our maps has elevation information: the curves indicate that how high is every location on the road and around. The first variable is time, t, again. We also have *spatial* variables: the horizontal location, points (x, y) on the plane and the elevation (the vertical location) z. Then z depends on (x, y) and (x, y) depends on t. Therefore, z depends on t via the *composition*. We can take specific functions:

- The horizontal location as a function of time, $x = 2t$ and $y = \sin t$;
- The elevation as a function of horizontal location, $z = x^2 + y^2$.
- Then, the elevation as a function of time, $z = (2t)^2 + (\sin t)^2$.

We can now answer the question directly:

$$\frac{dz}{dt} = \frac{d}{dt}((2t)^2 + (\sin t)^2) = 8t + 2\sin t \cos t\,.$$

Is there the Chain Rule? If there is, the above derivative is made of the following four. First, our motion is recorded as a parametric curve and its derivative consists of the derivatives of the two components:

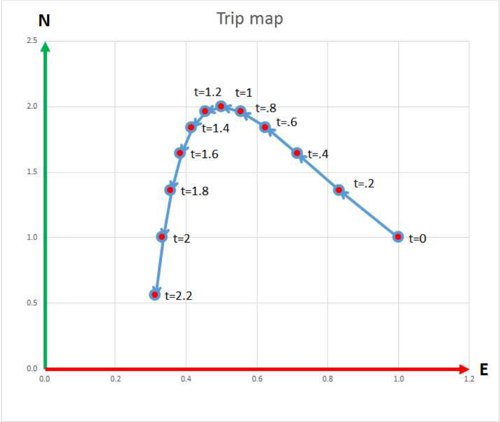
$$\frac{dx}{dt} = 2 \quad \text{and} \quad \frac{dy}{dt} = \cos t\,.$$

The terrain map's steepness is found in the two main direction, x and y, as the two partial derivatives of the function:

$$\frac{\partial z}{\partial x} = 2x \quad \text{and} \quad \frac{\partial z}{\partial y} = 2y\,.$$

Example 1.1.3: hiking

Let's now consider a more complex trip. Planning a hike, we create a *trip plan*: The times and the places are put on a simple map of the area:



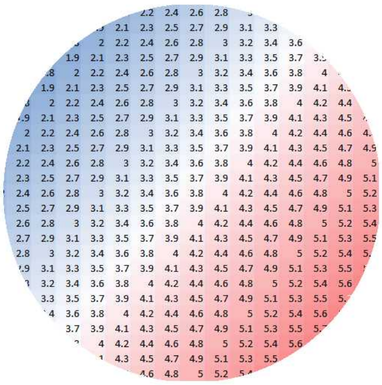
This is a *parametric curve*:

$$x = f(t),\ y = g(t)\,,$$

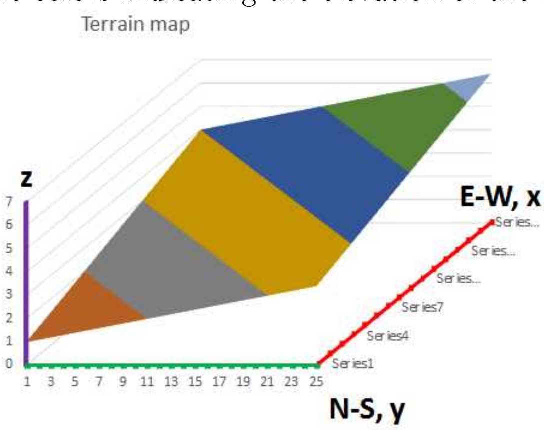
with x and y providing the coordinates of your location. Conversely, motion in time is a go-to metaphor for parametric curves!

We then bring the *terrain map* of the area. The data is colored accordingly:

y\x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5
0	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4
0.1	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
0.2	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2
0.3	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3
0.4	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4
0.5	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5
0.6	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6
0.7	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7
0.8	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8
0.9	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9
1	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5
1.1	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1
1.2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2
1.3	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3
1.4	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4
1.5	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5
1.6	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4	5.6
1.7	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5	5.7
1.8	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4	5.6	5.8
1.9	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.9
2	3	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4	5.6	5.8	6
2.1	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.9	6.1
2.2	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4	5.6	5.8	6	6.2
2.3	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.9	6.1	6.3
2.4	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5	5.2	5.4	5.6	5.8	6	6.2	6.4



Such a topographic map has the colors indicating the elevation of the actual terrain:

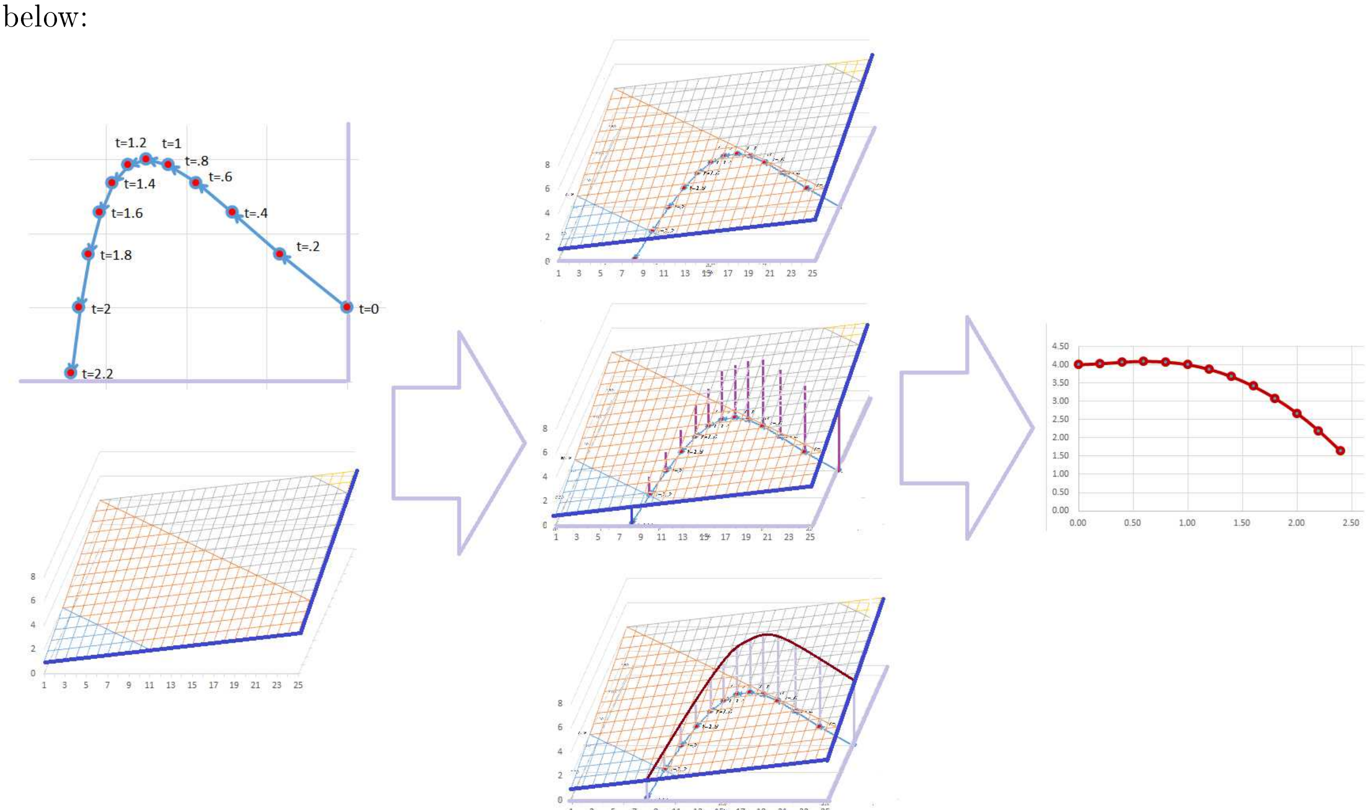


This is a *function of two variables*:

$$z = f(x, y)\,.$$

Conversely, a terrain map is a go-to metaphor for functions of two variables!

Now, back to the same question: How fast will we be climbing? The composition required is illustrated below:



We face new kinds of functions:

trip map				
t	$\longrightarrow$	(x,y)	$\longrightarrow$	z
		terrain map		

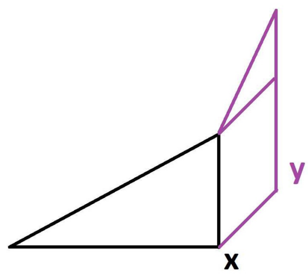
Both functions deal with 3 variables at the same time, with a total of 4!

In the meantime, there are many functions of the 2 or 3 variables of location: the temperature and the pressure of the air or water, the humidity, the concentration of a particular chemical, etc.

The observations about the rates of change are still applicable:

- If we double our horizontal speed (with the same terrain), the climb will be twice as fast.
- If we double steepness of the terrain (with the horizontal speed), the climb will be twice as fast.

It follows that the speed of the climb is proportional to both our horizontal speed and the steepness of the terrain. That's the Chain Rule.



What is it in this new setting? Both rates of change of and with respect to x and y will have to be involved. We will show that it is the sum of those:

$$\frac{\Delta z}{\Delta t} = \frac{\Delta z}{\Delta x}\frac{\Delta x}{\Delta t} + \frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta t} \quad \text{or} \quad \frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\,.$$

This number is then computed from these two *vectors*:

- the rate of change of the parametric curve of the trip, i.e., the horizontal velocity:

$$\left\langle \frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t} \right\rangle \quad \text{or} \quad \left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle\,.$$

- The rates of change of the terrain function in the two directions:

$$\left\langle \frac{\Delta z}{\Delta x}, \frac{\Delta z}{\Delta x} \right\rangle \quad \text{or} \quad \left\langle \frac{\partial z}{\partial x}, \frac{\partial z}{\partial x} \right\rangle\,.$$

Example 1.1.4: costs and prices

We saw in Volume 3 (Chapter 3IC-4) an example of an abstract space: the space of prices. At its simplest, the baker does two things:

1. He watches the prices of the two ingredients of his bread: sugar and wheat.
2. He decides, based on these two numbers, what the price of the bread should be.

The space's dimension was 2, with only the two prices of the two ingredients of bread. The dependence

is just as in the last example:

costs				
t	$\mapsto$	(x,y)	$\mapsto$	z
		price		

Multiple variables lead to high-dimensional abstract spaces, such as in the case of the price of a car dependent on the prices of 1000 of its parts:

$$t \ \mapsto \ (x_1, x_2, ..., x_{1000}) \ \mapsto \ z$$

We can develop algebra, geometry, and calculus that will be applicable to a space of *any* dimension. We replace a large number of single variables with a single variable in a space of a large dimension. For example, P below is such a variable, i.e., a location in a 1000-dimensional space:

$$t \ \mapsto \ P \ \mapsto \ z$$

Initially however we will limit ourselves to dimensions that we can visualize!

CONVENTION

We will use *upper case* letters for the entities that are (or may be) multidimensional, such as points and vectors:

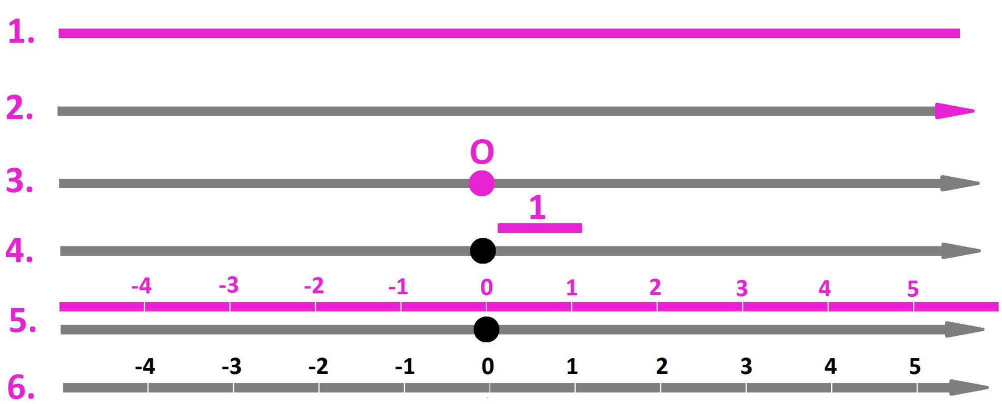
$$A,\ B,\ C,\ ...\ P,\ Q,\ ...,$$

and *lower case* letters for numbers:

$$a,\ b,\ c,\ ...\ ,x,\ y,\ z,\ \ldots$$

1.2. *Euclidean spaces and Cartesian systems of dimensions* 1, 2, 3,...

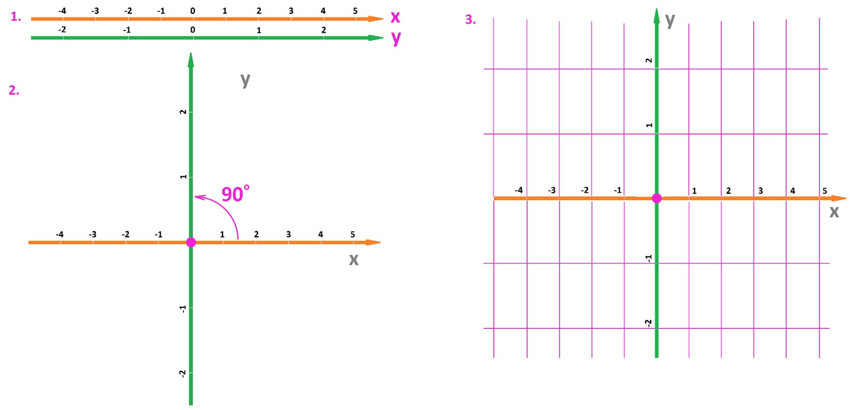
We start with the Cartesian system for dimension 1. It is a line with a certain collection of features – the origin, the positive direction, and the unit – added:



The main idea is this correspondence (i.e., a function that is one-to-one and onto):

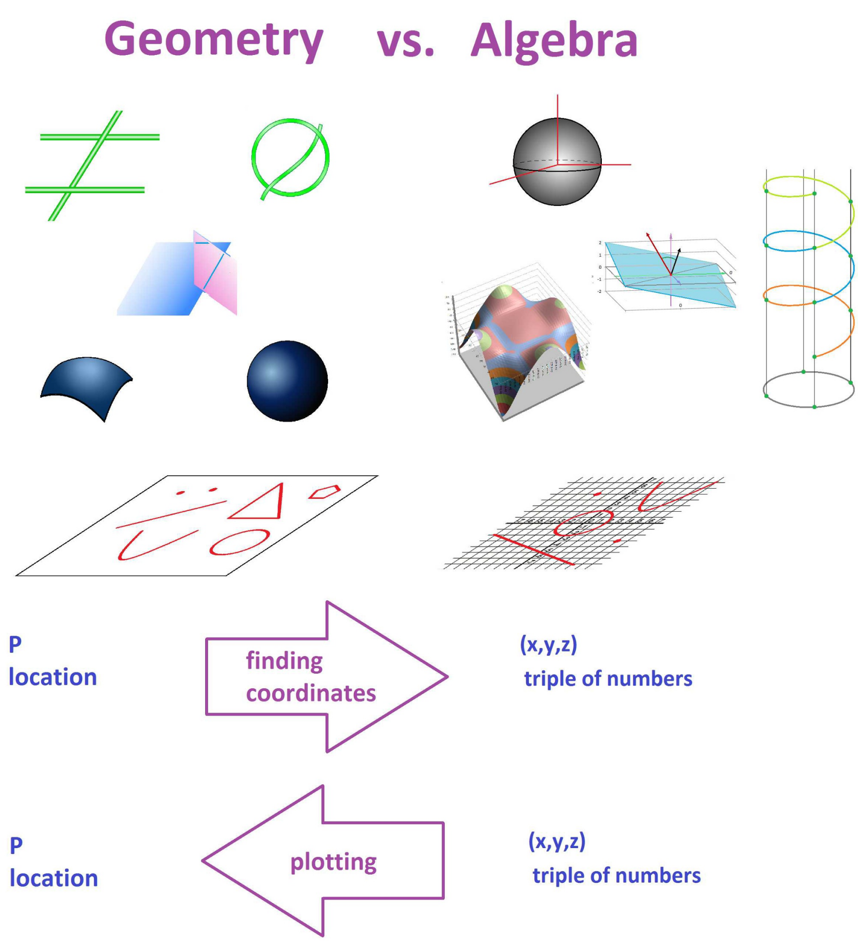
$$\boxed{\text{a location } P \longleftrightarrow \text{ a real number } x.}$$

We can have such "Cartesian lines" as many as we like and we can arrange them in any way we like. Then the Cartesian system for dimension 2 is made of *two* copies of the Cartesian system of dimension 1 aligned at 90 degrees (of rotation) from positive x to positive y:



We continue with *dimension* 3.

There is much more going on in "space" than on a plane:



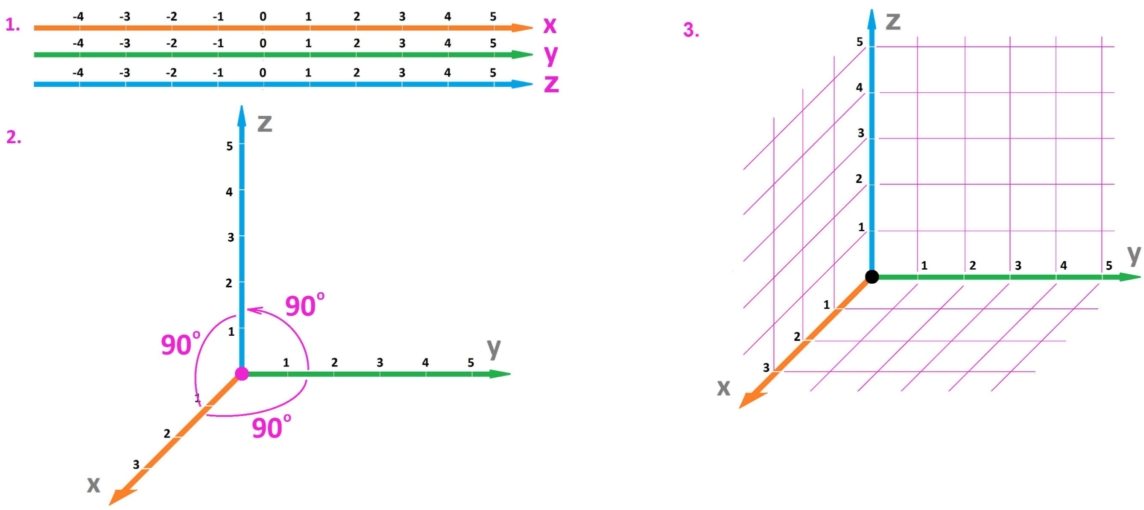
That is why we'll need *three* numbers to represent the locations.

The Cartesian system for dimension 3 is made of *three* copies of the Cartesian system of dimension 1. Just like in the case of dimension 2 above, these copies don't have to be identical; their units might be unrelated.

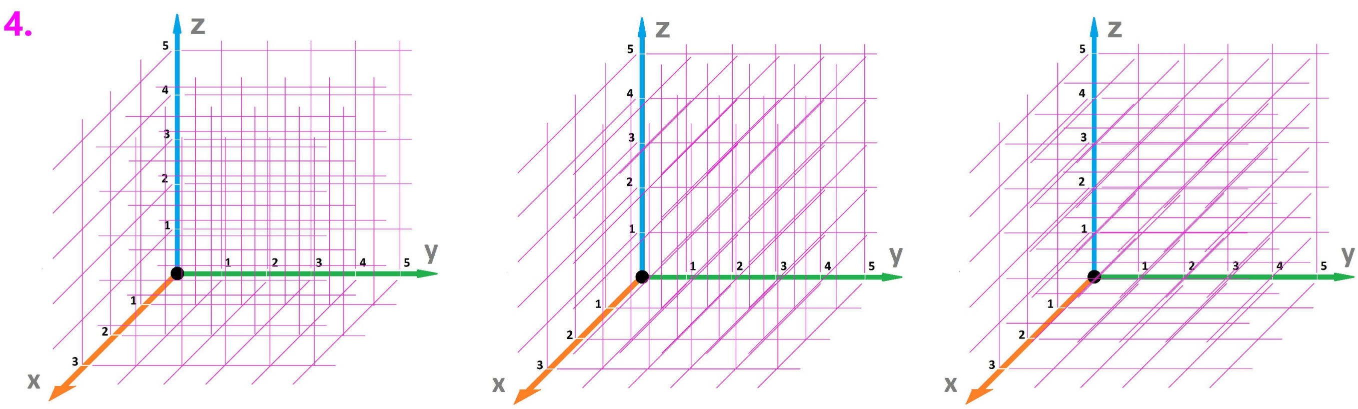
The system is built in several stages:

1. Three *coordinate axes* are chosen: the x-axis, the y-axis, and the z-axis.
2. The two axes are put together at their origins so that it is a 90-degree turn from the positive direction of one axis to the positive direction of the next – from x to y to z to x.
3. Use the marks on the axis to draw grids on the planes.

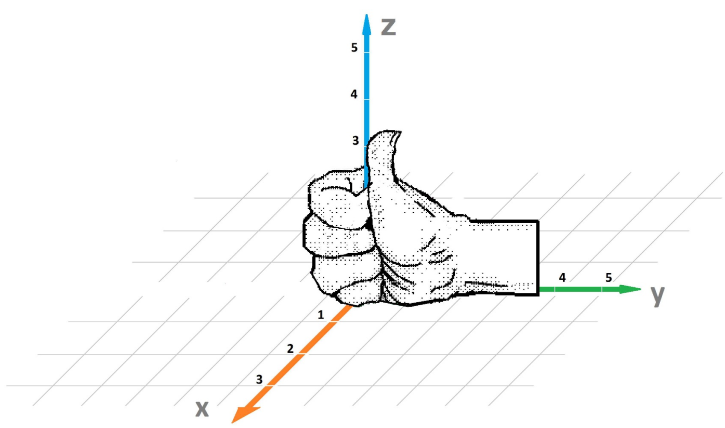
4. We repeat these three grids in parallel to create threads in space.



The last step is shown below:



The second requirement is called the *Right Hand Rule*:

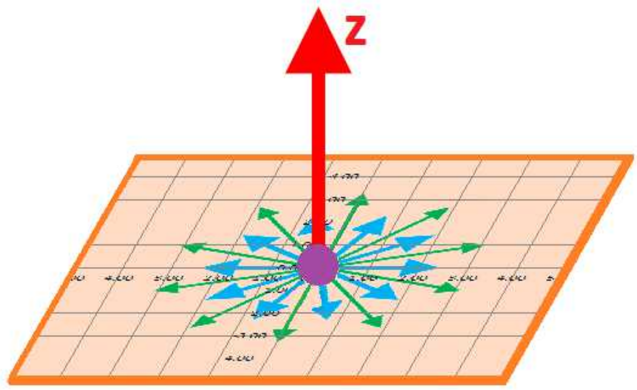


It reads:

► If we curl our fingers from the x-axis to the y-axis, our thumb will point in the direction of the z-axis.

We can also understand this idea if we imagine turning a screwdriver in this direction and seeing which way the screw goes.

The axes are perpendicular to each other, but there is more! For example, in addition to the x- and y-axis being perpendicular to the z-axis, *all* lines in the xy-plane are perpendicular to it:



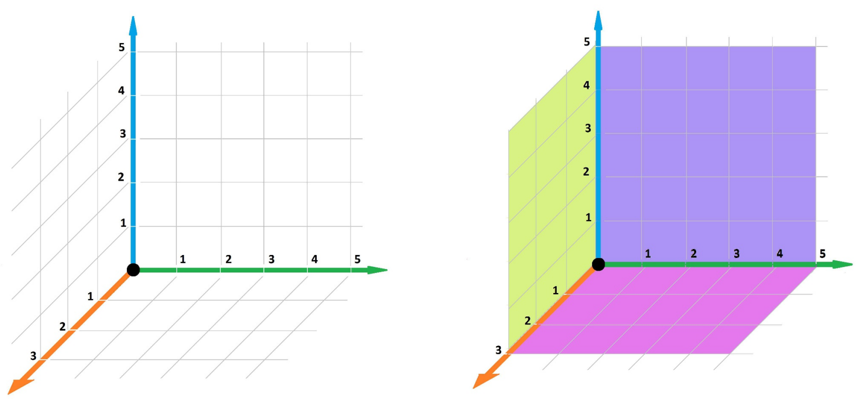
The main purpose of the Cartesian system remains the same; it is this correspondence:

$$\boxed{\text{a location } P \longleftrightarrow \textit{a triple} \text{ of real numbers } (x, y, z)}$$

Warning!

The three variables or quantities represented by the three axes may be unrelated. Then our visualization will remain valid with rectangles instead of squares, and boxes instead of cubes.

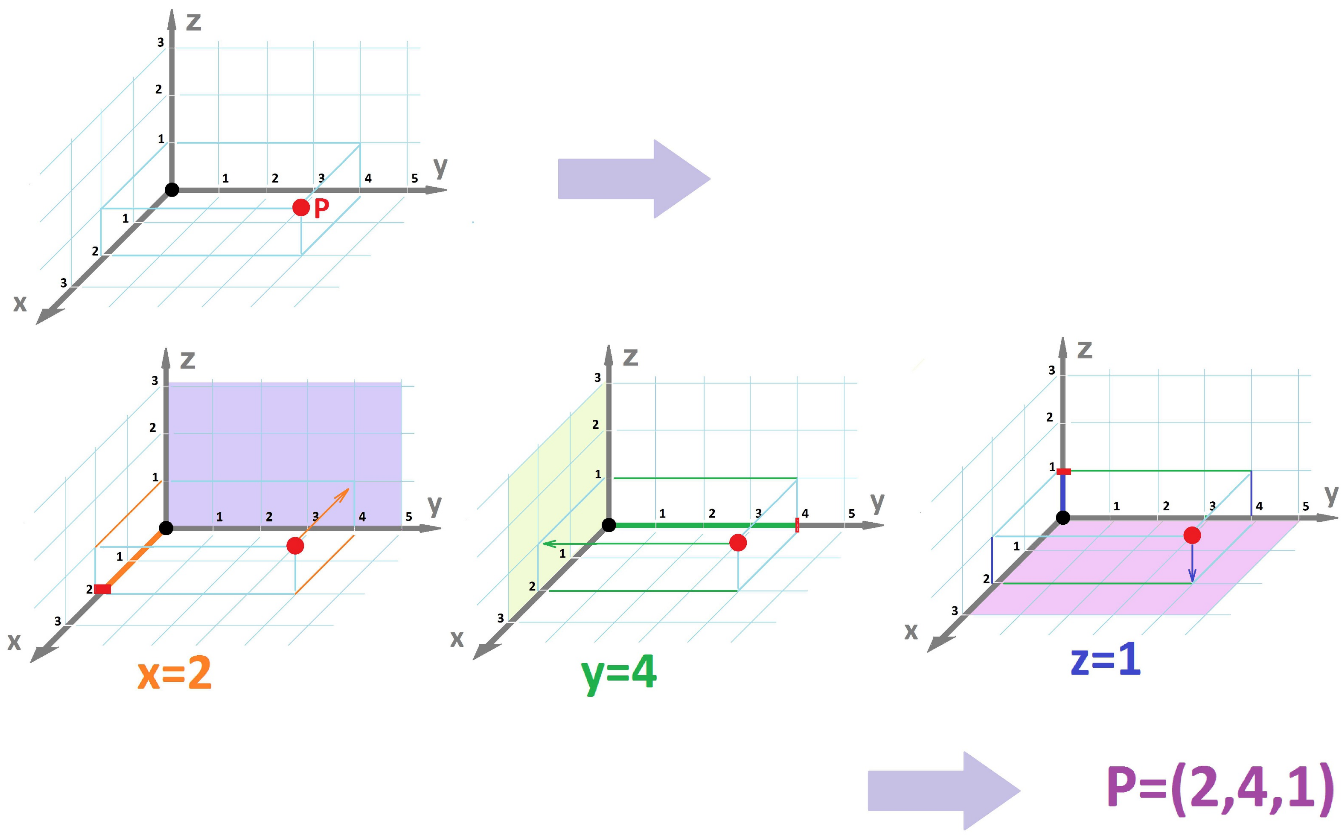
Alternatively, the system is built from three copies of the Cartesian plane: the xy-plane, the yz-plane, and the zx-plane. They are arranged at 90 degrees as walls of a room:



These planes are called the *coordinate planes*.

This is how the system works:

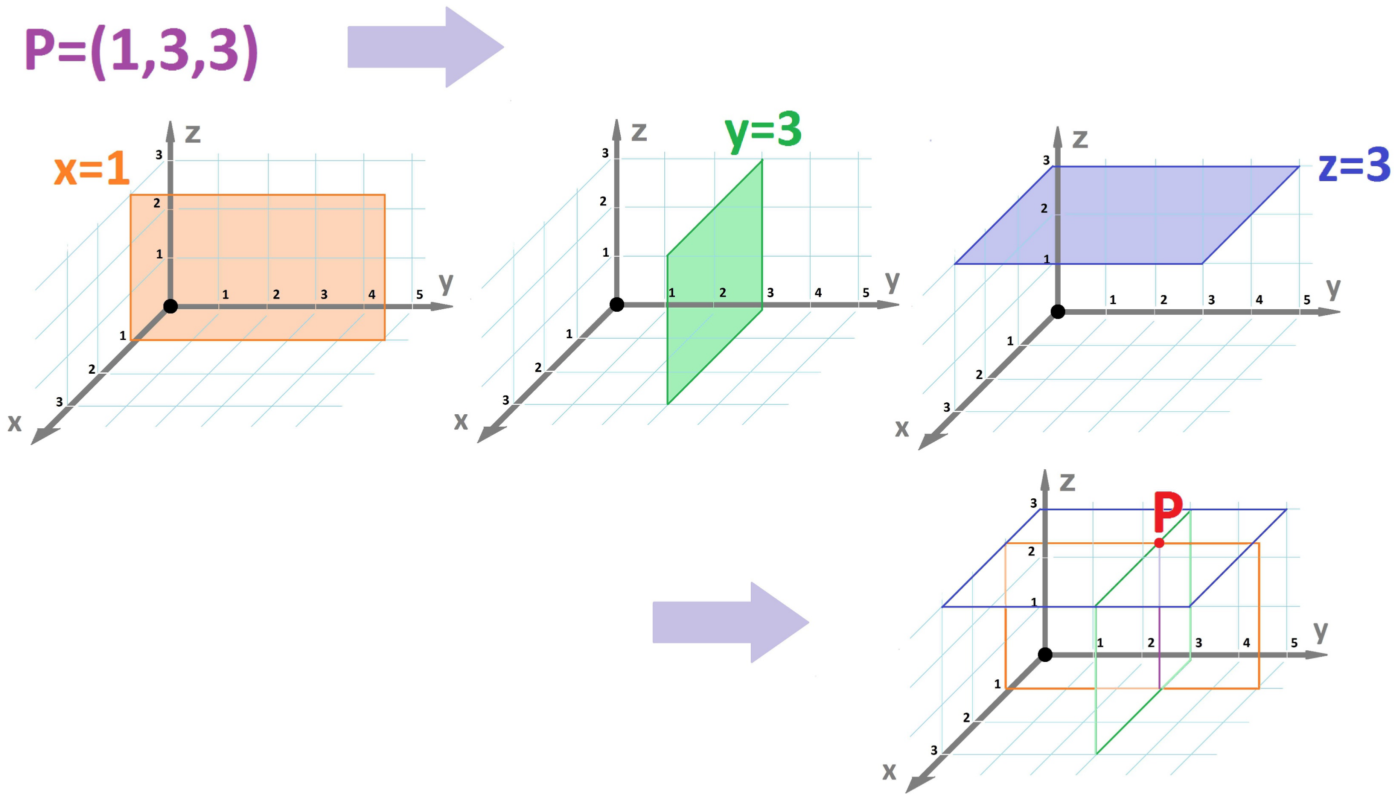
First, suppose P is a *location* in this space. We find the dimensions of the *box* with one corner at O and the opposite at P. We find the distances from the three planes to that location – positive in the positive direction and negative in the negative direction – and the result is the three coordinates of P, some *numbers* x, y, and z. The distance from the yz-plane is measured along the x-axis, etc. We use the nearest mark to simplify the task:



Conversely, suppose x, y, z are *numbers*. If we need to build a box with these dimensions:

- First, we measure x as the distance from the yz-plane – positive in the positive direction and negative in the negative direction – along the x-axis and create a plane parallel to the yz-plane.
- Second, we measure y as the distance from the xz-plane along the y-axis and create a plane parallel to the xz-plane.
- Third, we measure z as the distance from the xy-plane along the z-axis and create a plane parallel to the xy-plane.

The intersection of these three planes – as if these were the two walls and the floor in a room – is a *location* $P = (x, y, z)$ in the space. We use the nearest marks to simplify the task:

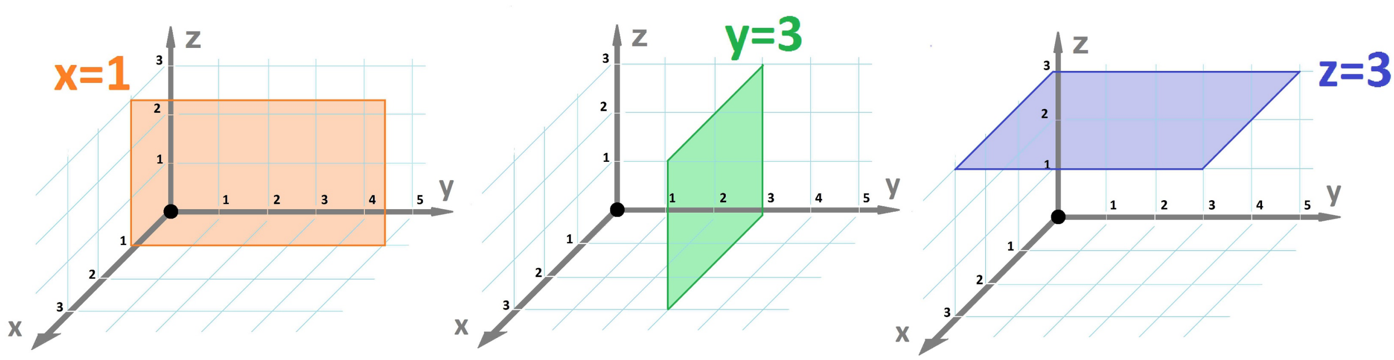


This 3-dimensional coordinate system is called *the Cartesian space* or the 3-*space*.

Once the coordinate system is in place, it is acceptable to think of location as triples of numbers and vice versa. In fact, we can write:

$$P = (x, y, z)\,.$$

Consider more of the planes parallel to the coordinate planes:



Then, we have a compact way to represent these planes:

$$x = k,\ \ y = k, \quad \text{or } z = k\,,$$

for some real k.

We can use this idea to reveal the internal structure of the space.

Theorem 1.2.1: Planes Parallel to Coordinate Planes

1. *If L is a plane parallel to the xy-plane, then all points on L have the same z-coordinate. Conversely, if a collection L of points consists of all points with the same z-coordinate, L is a plane parallel to the xy-plane.*
2. *If L is a plane parallel to the yz-plane, then all points on L have the same x-coordinate. Conversely, if a collection L of points consists of all points with the same x-coordinate, L is a plane parallel to the yz-plane.*

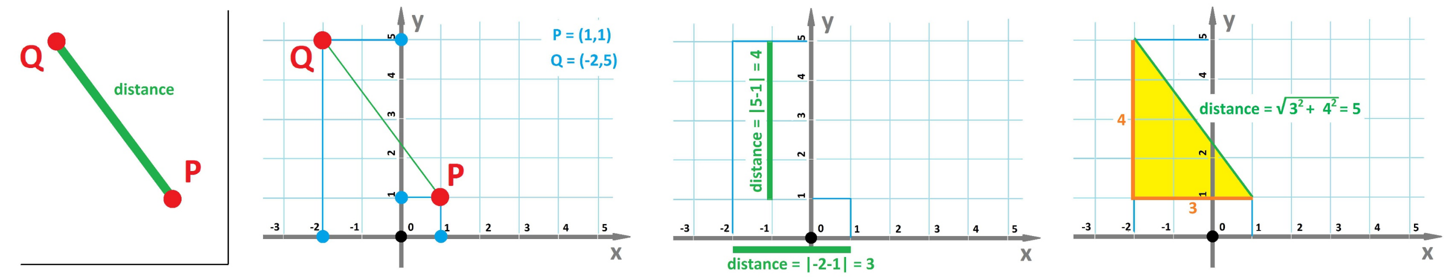
> 3. *If L is a plane parallel to the zx-plane, then all points on L have the same y-coordinate. Conversely, if a collection L of points consists of all points with the same y-coordinate, L is a plane parallel to the zx-plane.*

We turn to *analytic geometry* of the 3-space.

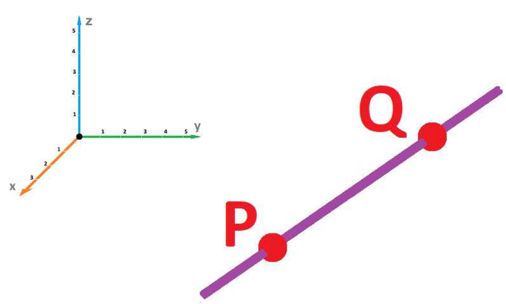
Now that everything is *pre-measured*, we can solve geometric problems by algebraically manipulating coordinates.

The first geometric task is finding the *distance*: What is the distance between locations P and Q in terms of their coordinates (x, y, z) and (x', y', z')?

For dimension 2, we used the distance formula from the 1-dimensional case. We found distance between two points on the plane as the length of the diagonal of the rectangle – with its sides parallel to the coordinate axes – that has these points at the opposite corners:

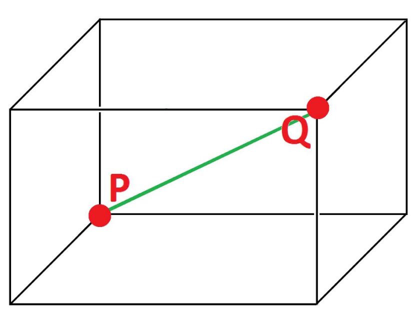


First, we need to realize that the problem itself is 1-dimensional! Indeed, any two points, in any space – 1-, 2-, 3-, or n-dimensional – can be connected by a line, and along that line – a 1-dimensional space – we measure the distance:



The coordinate system is just a means to an end.

So, we need to find the distance between two points in space as the length of the diagonal of the *box* – with its edges parallel to the coordinate axes, and sides parallel to the coordinate planes – that has these points at the opposite corners:

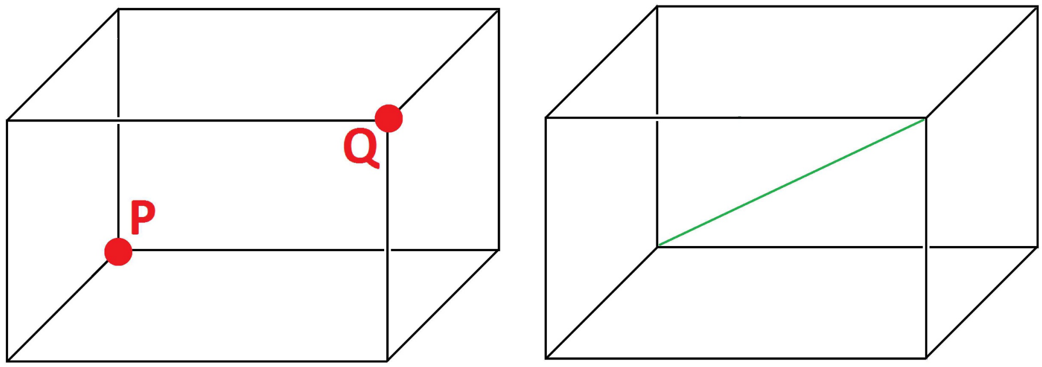


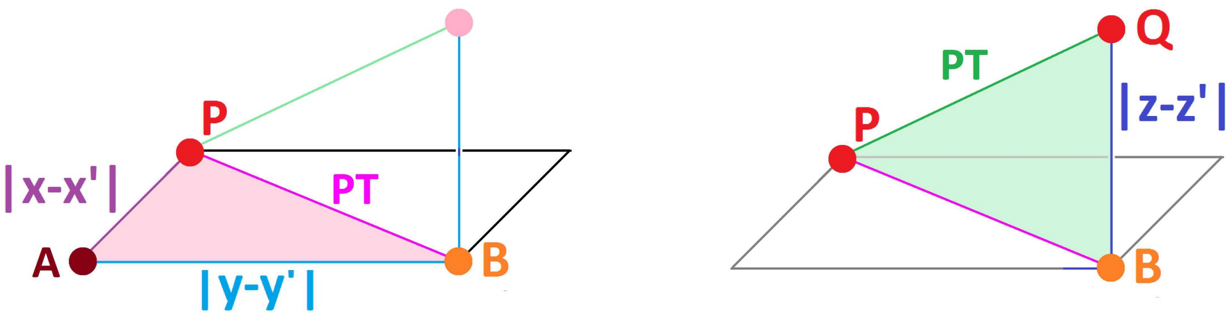
We now utilize these two facts:

1. Every coordinate plane of the 3-space has its own, 2-dimensional, coordinate system.

2. The coordinate axes are perpendicular to the coordinate planes.

This is the outline of the construction:





The Pythagorean theorem is to be applied within the horizontal plane and then within a certain vertical plane.

The formula is, as we anticipated, symmetric with respect to the dimensions:

Theorem 1.2.2: Distance Formula For Dimension 3

The distance between points with coordinates $P = (x, y, z)$ and $Q = (x', y', z')$ is

$$\boxed{d(P, Q) = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}}$$

Proof.

The segment between the points $P = (x, y, z)$ and $Q = (x', y', z')$ is the diagonal of this "box". We use the distance formula from the 1-dimensional case separately for each of the three axes, as follows:

1. The distance between x and x' on the x-axis is $|x - x'|$.
2. The distance between y and y' on the y-axis is $|y - y'|$.
3. The distance between z and z' on the z-axis is $|z - z'|$.

These are the dimensions of the box.

Next we use the *Pythagorean Theorem* twice. We first find the length of the diagonal of the bottom of the box and then the length of the main diagonal:

$$\begin{array}{llll}
\text{PT 1:} & d(P, A) = |x - x'|, & d(A, B) = |y - y'| & \\
& & \Longrightarrow \quad d(P, B)^2 & = (x - x')^2 + (y - y')^2 \\
\text{PT 2:} & d(P, B)^2 = (x - x')^2 + (y - y')^2, & d(B, Q) = |z - z'| & \\
& & \Longrightarrow \quad d(P, Q)^2 & = d(P, B)^2 + d(B, Q)^2 \\
& & & = (x - x')^2 + (y - y')^2 + (z - z')^2
\end{array}$$

Exercise 1.2.3

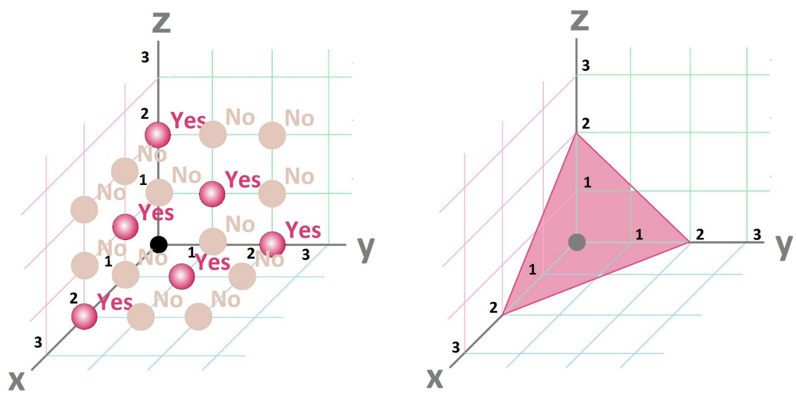
Prove that in the latter case the triangle is indeed a right triangle.

Relations are used in the same way as before but with more variables. A relation processes a triple of numbers (x, y, z) as the input and produces an output, which is: Yes or No. If we are to plot the *graph* of

a relation, this output becomes: a point or no point. For example:

plane:		triple:		relation:			outcome:
						`TRUE` $\nearrow$	Plot point (x,y,z).
$\mathbf{R}^3$	$\to$	(x,y,z)	$\to$	$\boxed{x+y+z=2\,?}$	$\to$		
						`FALSE` $\searrow$	Don't plot anything.

We can do it by hand:



We can use, as before, the *set-building notation*:

$$\{(x,y,z):\quad \text{a condition on } x,y,z\}\,.$$

For example, the graph of the above relation is a subset of $\mathbf{R}^3$ given by:

$$\{(x,y,z):\; x+y+z=2\}\,.$$

What about dimension 4 and higher?

We cannot use our physical space as a reference anymore! We can't use it for visualization either. The space is abstract.

The idea of the n-dimensional space remains the same; it is the correspondence:

$$\boxed{\text{a location } P \;\longleftrightarrow\; \text{a string of } n \text{ real numbers } (x_1,x_2,x_3,...,x_n)}$$

Using the same letter with subscripts is preferable even for dimension 3 as the symmetries between the axes and variables are easier to detect and utilize. However, using just P is often even better!

Because of the difficulty or even impossibility of visualization of these "locations" in dimension 4, this correspondence becomes much more than just a way to go back and forth whenever convenient. This time, we just say "It's the same thing".

Example 1.2.4: non-homogeneous variables

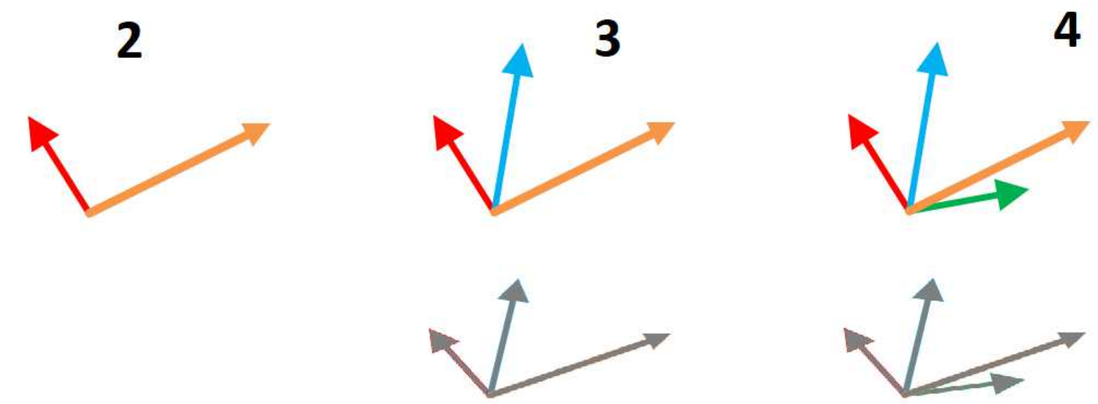
This may be the data continuously collected by a *weather center*:

1	2	3	4	5	...
temperature	pressure	precipitation	humidity	sunlight	...

They are all measured in different units and cannot be seen as an analog of our physical space.

How do we visualize this n-dimensional space?

Let's first realize that, in a sense, we have failed even with the *three*-dimensional space! We have had to squeeze these three dimensions on a *two*-dimensional piece of paper. Without the numbers telling us what to expect, we wouldn't be able to tell the dimension (top row):



At best, we are seeing the *shadows* of the lines (bottom row).

These are the spaces we will study and the notations for them:

Euclidean spaces

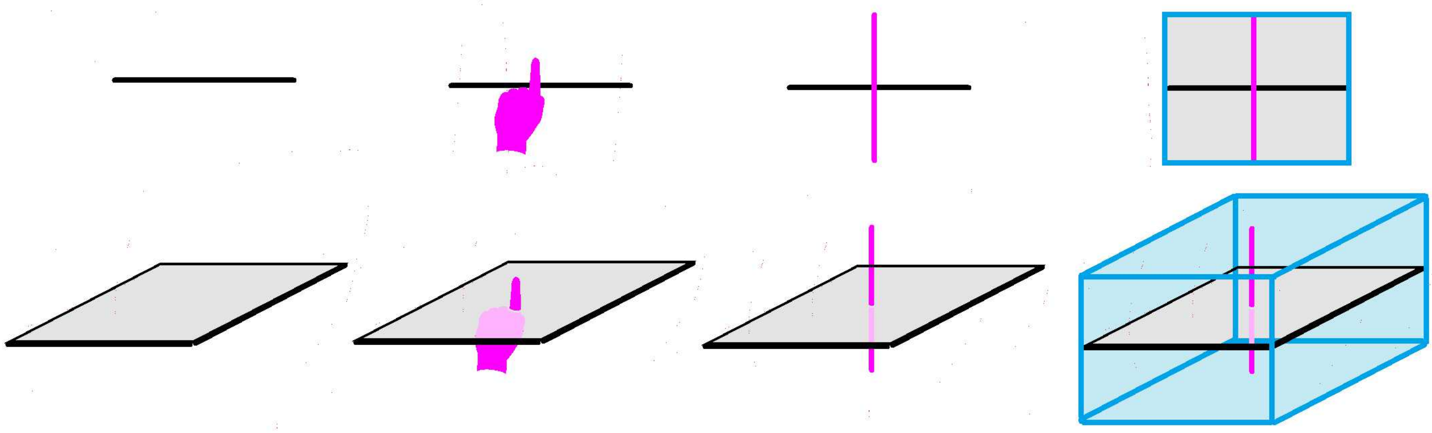
- $\mathbf{R}$, all real numbers (line)
- $\mathbf{R}^2$, all pairs of real numbers (plane)
- $\mathbf{R}^3$, all triples of real numbers (space)
- $\mathbf{R}^4$, all quadruples of real numbers
- ...
- $\mathbf{R}^n$, all strings of n real numbers
- ...

Each of them is supplied with its own algebra and geometry.

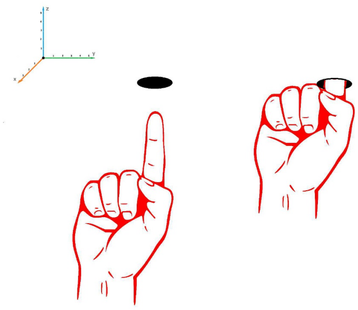
We can build these by consecutively adding one dimension at a time.

- If $\mathbf{R}$ is given, we treat it as the x-axis and then add another axis, the y-axis, perpendicular to the first.
- The result is $\mathbf{R}^2$, which we treat as the xy-plane and then add another axis, the z-axis, perpendicular to the first two.
- The result is $\mathbf{R}^3$, which we treat as the xyz-space and then add another axis perpendicular to the first three; and so on.

Here is the summary:



With our 1-dimensional finger, we puncture the space. In $\mathbf{R}^4$, the same thing happens; the finger disappears:



The formula that represents the line in the first row is:

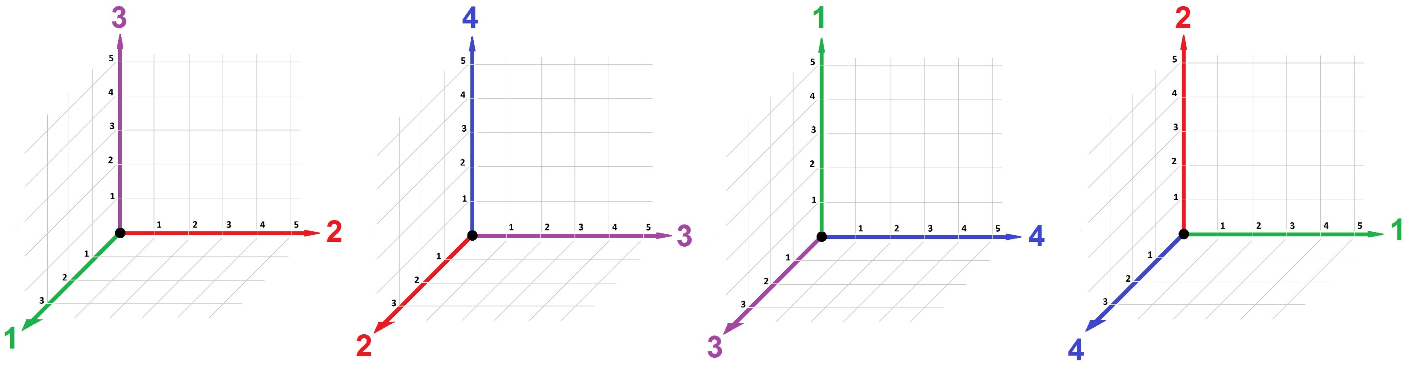
$$y = 0 \text{ or } x_2 = 0\,.$$

The formula that represents the plane in the second row is:

$$z = 0 \text{ or } x_3 = 0\,.$$

This space is abstract but is still constructed from lower-dimensional spaces:

1. Four copies of $\mathbf{R}$: the four coordinate axes.
2. Six copies of $\mathbf{R}^2$: the six coordinate planes, each spanned on a pair of those coordinate axes.
3. Four copies of $\mathbf{R}^3$: four spaces, each constructed on the frame of three of those coordinate planes.



Exercise 1.2.5

What is the formula that represents each of these spaces?

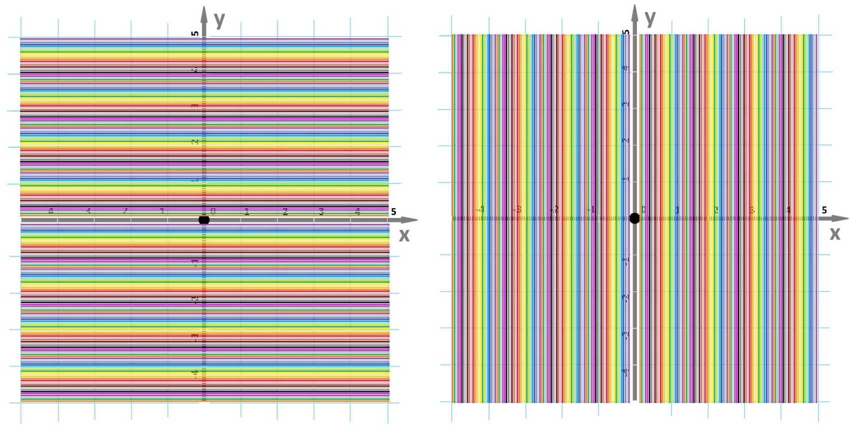
Exercise 1.2.6

How many coordinate planes are there in $\mathbf{R}^5$? $\mathbf{R}^n$? How many coordinate spaces?

So, these spaces aren't unrelated!

In order to reveal the internal structure of a spaces, we look for *lower-dimensional* spaces in it.

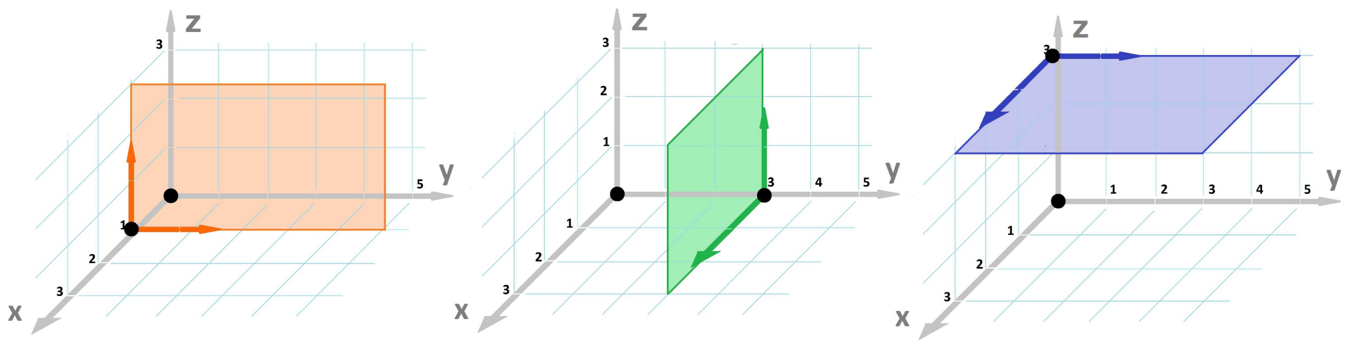
The plane is a *stack of lines*, each of which is just a copy of one of the coordinate axes:



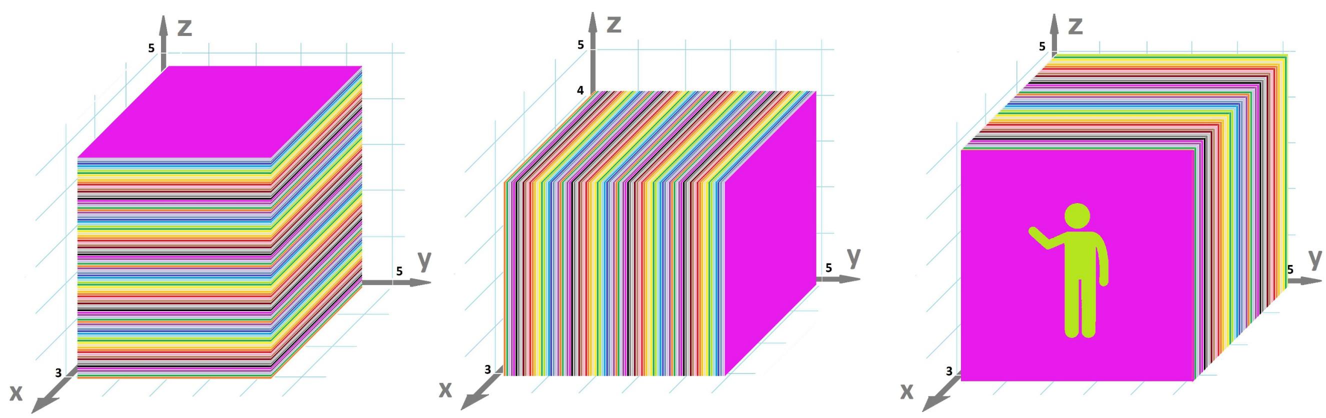
These lines are given by the equations for each real a or b:

$$\boxed{x = a,\ y = b}$$

They are *copies* of $\mathbf{R}$ and will have the same algebra and geometry. In fact, they can have their own coordinate systems:



Now, one can think of the 3-space as a *stack of planes*, each of which is just a copy of one of the coordinate planes:



They are given by the equations for all real a, b, c:

$$\boxed{x = a,\ y = b,\ z = c}$$

These are copies of $\mathbf{R}^2$.

If a 2-dimensional person can recognize – thinking mathematically – that the 3-space is made of layers of copies of his own space, we can see our physical 3-space as just a single "layer" in $\mathbf{R}^4$.

So, $\mathbf{R}^4$ is a "stack" of $\mathbf{R}^3$s. How they fit together is hard to visualize, but they are still copies of $\mathbf{R}^3$ given by equations:

$$\boxed{x_1 = a_1,\ x_2 = a_2,\ x_3 = a_3,\ x_4 = a_4}$$

Exercise 1.2.7

What is a line a stack of?

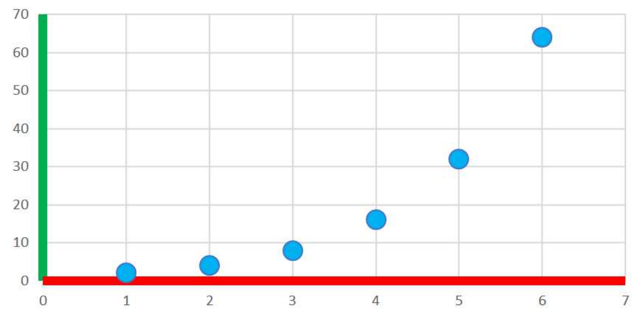
So, we can see many copies of $\mathbf{R}^m$ in $\mathbf{R}^n$, with $n > m$.

Beyond a certain point, the chance to visualize the space is gone. We, however, are still able to visualize the space one element at a time. For example, a point in the n-dimensional space is nothing but a *sequence* with n terms:

k	1	2	3	4	5	...	n
x_k	x_1	x_2	x_3	x_4	x_5	...	x_n

It's just a function, with the inputs in the first row and the outputs in the second. We visualize functions with their *graphs*. For example, this string of 6 numbers is a point in the 6-dimensional space:

n	y
1	2.0
2	4.0
3	8.0
4	16.0
5	32.0
6	64.0

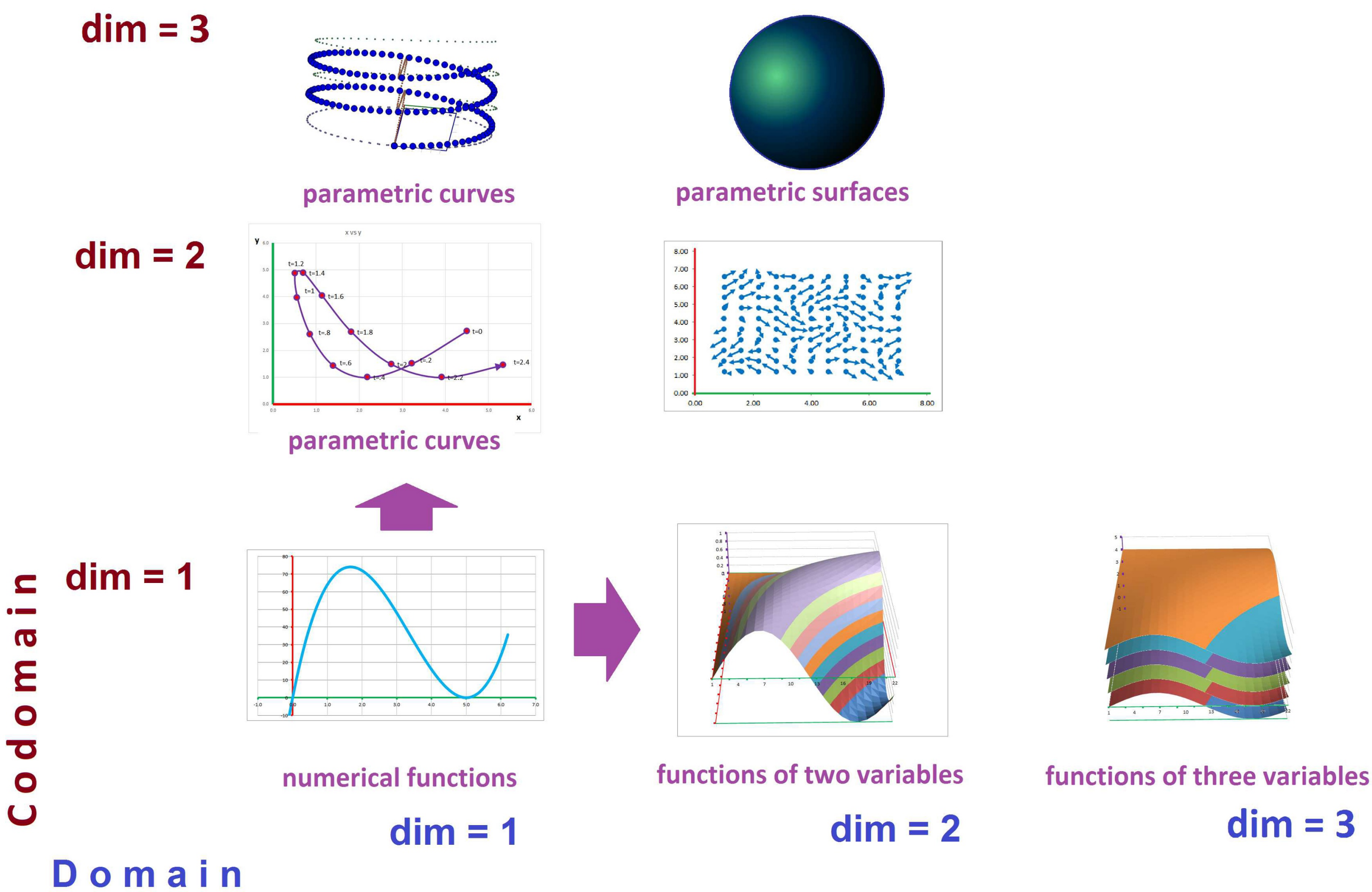


Warning!

The curve that you see is incidental because the rows of the table can be re-arranged.

Next, there is no point in studying spaces without studying *functions* between them.

Let's review multidimensional functions. We place them in a table with two axes representing the dimension of the domain and the dimension of the codomain:



We always start at the very first cell. Previously we made a step in the vertical direction and explored the first column of this table. We also moved to the right.

With all this complexity, we shouldn't overlook the general point of view on functions. We represent a function diagrammatically as a *black box* that processes the input and produces the output of whatever nature:

$$\begin{array}{ccccc} \text{input} & & \text{function} & & \text{output} \\ X & \mapsto & \boxed{f} & \mapsto & Y \end{array}$$

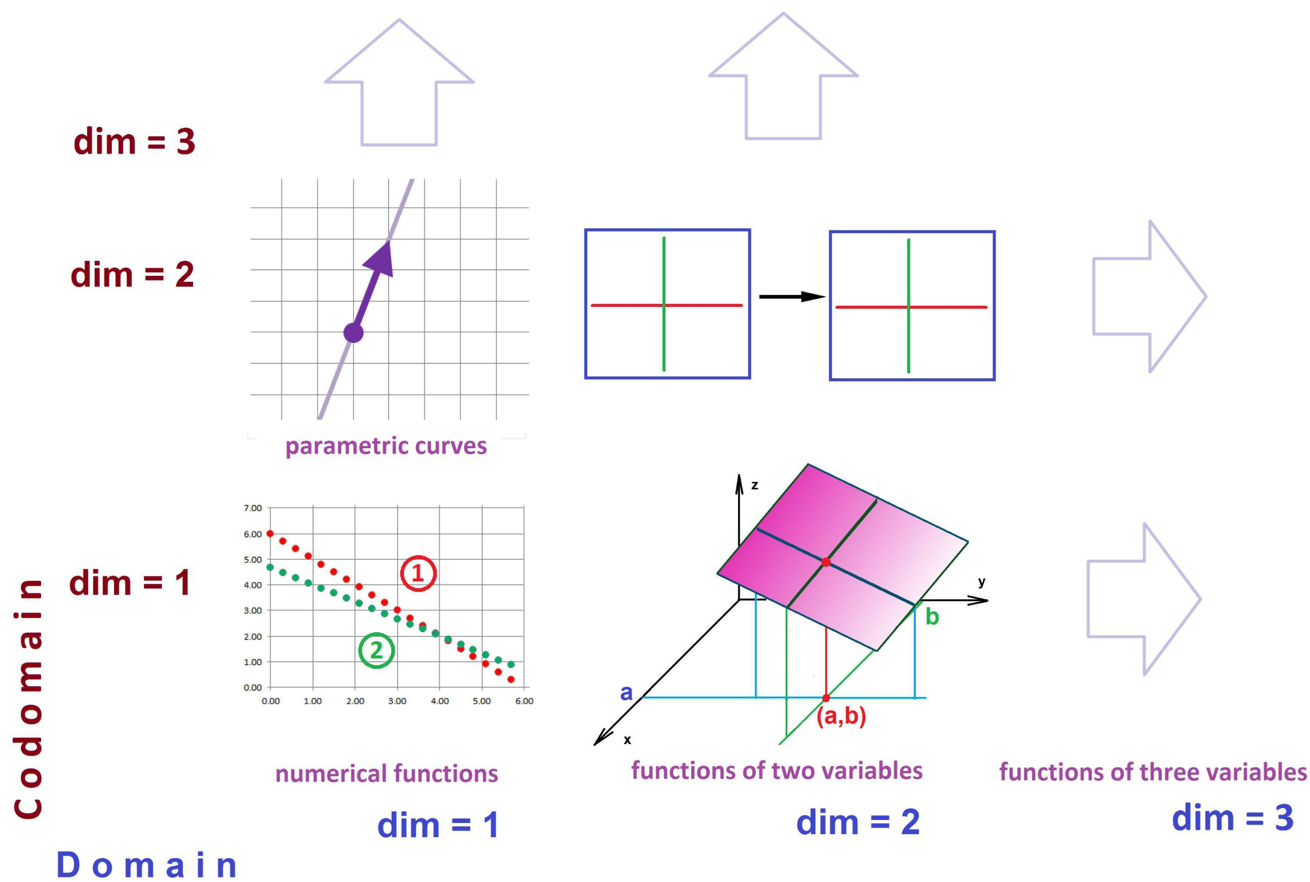
Let's take one from the left-most column and one from the bottom row:

input		parametric curve		output
t	$\mapsto$	$\boxed{F}$	$\mapsto$	X
$\mathbf{R}$				$\mathbf{R}^m$
number				point
time				prices of parts
time				prices of stocks

input		function of two variables		output
X	$\mapsto$	$\boxed{f}$	$\mapsto$	z
$\mathbf{R}^m$				$\mathbf{R}$
point				number
prices of parts				price of car
prices of stocks				value of portfolio

They can be linked up and produce a composition, which is just a numerical function.

Above is a view of "generic" functions. In the linear algebra context, the functions in the table are simpler and so are their visualizations:



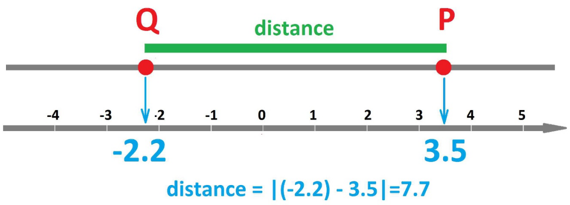
We turn to analytic geometry of n-dimensional spaces.

1.3. *Geometry of distances*

The axes of the Cartesian system $\mathbf{R}^3$ for our physical space refer to the same: distances to the coordinate planes. They are (or should be) measured in the same unit. Even though, in general, the axes of $\mathbf{R}^n$ refer to unrelated quantities, they *may* be measured in the same unit, such as the prices of n commodities being traded. When this is the case, doing geometry in $\mathbf{R}^n$ based entirely on the coordinates of points is possible.

A Cartesian system has everything in the space *pre-measured*.

In particular, we compute (rather than measure) the distances between locations because the distance can be expressed in terms of the coordinates of the locations.



Theorem 1.3.1: Distance Formula for Dimension 1

The distance from point P to point Q in $\mathbf{R}$ given by real numbers x and x' respectively is

$$\boxed{d(P,Q) = |x - x'|}$$

Here, the geometry problem of finding distances relies on the algebra of real numbers (the subtraction).

Now the coordinate system for dimension 2. The formula for the distance between locations P and Q in terms of their coordinates (x, y) and (x', y') is found by using the distance formula from the 1-dimensional case for either of the two axes in order to find

1. the distance between x and x', which is $|x - x'| = |x' - x|$, and
2. the distance between y and y', which is $|y - y'| = |y' - y|$, respectively.

Then the two numbers are put together by the *Pythagorean Theorem* taking into account this simplification:

$$|x - x'|^2 = (x - x')^2, \ |y - y'|^2 = (y - y')^2 .$$

Theorem 1.3.2: Distance Formula for Dimension 2

The distance between points P and Q in $\mathbf{R}^2$ with coordinates (x, y) and (x', y') respectively is

$$\boxed{d(P,Q) = \sqrt{(x - x')^2 + (y - y')^2}}$$

The two exceptional cases when P and Q lie on the same vertical or the same horizontal line (and the triangle "degenerates" into a segment) are treated separately.

Now the coordinate system for dimension 3. We can guess that there will be another term in the sum of the *Distance Formula.*

Theorem 1.3.3: Distance Formula for Dimension 3

The distance between points P and Q in $\mathbf{R}^3$ with coordinates (x, y, z) and (x', y', z') respectively is

$$\boxed{d(P,Q) = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}}$$

A pattern starts to appear:

► The square of the distance is the sum of the squares of the distances along each of the coordinates.

Thinking by analogy, we continue on to include the case of dimension 4:

dimension	points	coordinates	distance
1	P	x	
	Q	x'	$d(P,Q)^2 = (x-x')^2$
2	P	(x,y)	
	Q	(x',y')	$d(P,Q)^2 = (x-x')^2 + (y-y')^2$
3	P	(x,y,z)	
	Q	(x',y',z')	$d(P,Q)^2 = (x-x')^2 + (y-y')^2 + (z-z')^2$
4	P	(x_1,x_2,x_3,x_4)	
	Q	(x_1',x_2',x_3',x_4')	$d(P,Q)^2 = (x_1-x_1')^2 + (x_2-x_2')^2 + (x_3-x_3')^2 + (x_4-x_4')^2$
...	...	...	...

There are n terms in dimension n:

dimension	points	coordinates	distance
n	P	$(x_1,x_2,...,x_n)$	
	Q	$(x_1',x_2',...,x_n')$	$d(P,Q)^2 = (x_1-x_1')^2 + (x_2-x_2')^2 + ... + (x_n-x_n')^2$

Example 1.3.4: geometry of $\mathbf{R}^3$ in $\mathbf{R}^4$

The formula for $n=1,2,3$ is justified by what we know about the physical space. What about $n=4$ and above? Let's take a look at the copies of $\mathbf{R}^3$ that make up $\mathbf{R}^4$. One of them is given by $x_4 = a_4$ for some real number a_4. If we take any two points P,Q within it, the formula becomes:

$$\begin{array}{lll}d(P,Q) & = & \sqrt{(x_1-x_1')^2 + (x_2-x_2')^2 + (x_3-x_3')^2 + (a_4-a_4)^2}\\ & = & \sqrt{(x_1-x_1')^2 + (x_2-x_2')^2 + (x_3-x_3')^2}\,.\end{array}$$

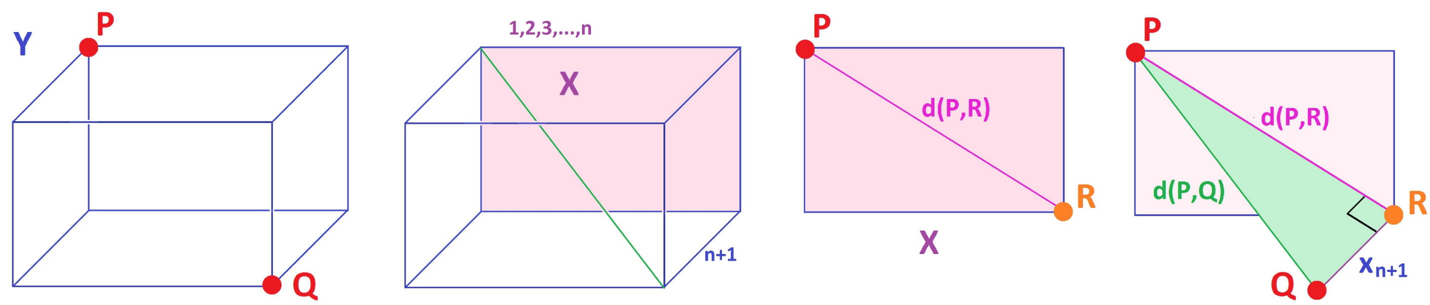
In other words, the distance is the same as the one for dimension 3. We conclude that the geometry of such a copy of $\mathbf{R}^3$ is the same as the "original"!

Exercise 1.3.5

Show that the geometry of *any* plane in $\mathbf{R}^3$ and $\mathbf{R}^4$ is the same as that of $\mathbf{R}^2$.

Can we justify this formula with more than just "It's a pattern"? Yes, we progress from understanding the geometry of $X = \mathbf{R}^n$ to that of $Y = \mathbf{R}^{n+1}$, every time.

Suppose the distances in $X = \mathbf{R}^n$ are computed by the above formula. Then, we add an extra axis – perpendicular to the rest – to create $Y = \mathbf{R}^{n+1}$:



Then, the *Pythagorean Theorem* is applied (the green triangle). The computation is just as in the case $n = 3$ presented above:

$$d(P,Q)^2 = d(P,R)^2 + d(R,Q)^2 = (x_1 - x_1')^2 + (x_2 - x_2')^2 + ... + (x_n - x_n')^2 + (x_{n+1} - x'_{n+1})^2 .$$

The formula applies to a space of any dimension n. It matches the measured distance in the physical space: $n = 1, 2, 3$. We can't say the same about the spaces of dimensions $n > 3$. They are abstract spaces. The formula, therefore, is seen as the *definition* of the distance for these spaces:

Definition 1.3.6: Euclidean metric

The *Euclidean distance between points* P and Q in $\mathbf{R}^n$ is defined to be the square root of the sum of the squares of the distances for each of the coordinates:

$$\begin{aligned} P &= (x_1, x_2, ..., x_n) \\ Q &= (x_1', x_2', ..., x_n') \\ d(P,Q) &= \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + ... + (x_n - x_n')^2} \end{aligned}$$

We refer to the formula as the *Euclidean metric*. The space $\mathbf{R}^n$ equipped with the Euclidean metric is called the *n-dimensional Euclidean space*.

Now, the formula is somewhat complicated. Is it possible to have a few simple rules apply equally to all dimensions, without reference to the formulas?

We formulate three very simple properties of the distances. First, the distances can't be negative and, moreover, for the distance to be zero, the two points have to be the same. Second, the distance from P to Q is the same as the distance from Q to P. And so on.

Theorem 1.3.7: Axioms of Metric Space

Suppose P, Q, S are points in $\mathbf{R}^3$. Then the following properties are satisfied:

- ***Positivity:*** *$d(P,Q) \geq 0$; and $d(P,Q) = 0$ if and only if $P = Q$.*
- ***Symmetry:*** *$d(P,Q) = d(Q,P)$.*
- ***Triangle Inequality:*** *$d(P,Q) + d(Q,S) \geq d(P,S)$.*

Proof.

Suppose $d(P,Q) = 0$. Then

$$0 = d(P,Q)^2 = (x_1 - x_1')^2 + (x_2 - x_2')^2 + ... + (x_n - x_n')^2 .$$

Since none of the terms is negative, all have to be zero:

$$(x_1 - x_1')^2 = 0,\ (x_2 - x_2')^2 = 0,\ ...,\ (x_n - x_n')^2 = 0 .$$

Therefore,

$$x_1 = x_1',\ x_2 = x_2',\ ...,\ x_n = x_n' .$$

It follows that $P = Q$.

Exercise 1.3.8

Prove the rest of the theorem.

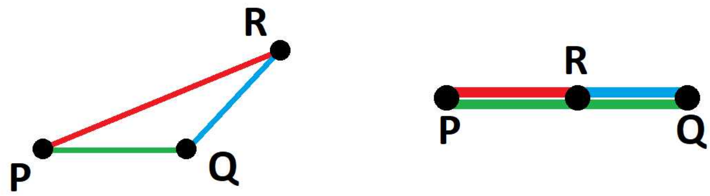
These geometric properties have been justified following the familiar geometry of the "physical space" $\mathbf{R}^3$.

However, they also serve as a *starting point* for further development of linear algebra. Below, we will define the new geometry of the abstract space $\mathbf{R}^n$ and demonstrate that these “axioms” are still satisfied.

Exercise 1.3.9

The distance is a function. Explain.

We know the last property from Euclidean geometry:



We can justify it for dimension $n \geq 4$ by referring to the following fact: Any three points lie within a single plane. This fact brings us back to Euclidean geometry... if that’s what we want.

Example 1.3.10: city blocks

The Distance Formula for the plane gives us the distance measured *along a straight line* as if we are walking through a field. But what if we are walking through a city? We then cannot go diagonally as we have to follow the grid of streets. This fact dictates how we measure distances. To find the distance between two locations $P = (x, y)$ and $Q = (u, v)$, we measure *along the grid* only:

The formula is, therefore:

$$\boxed{d_T(P, Q) = |x - u| + |y - v|}$$

It is called the *taxicab metric*. It is different from the Euclidean metric as the diagonal of a 1 square is 2 units long under this geometry.

Exercise 1.3.11

Prove that the taxicab metric satisfies the three properties in the theorem.

If the physical space can be reasonably treated with non-Euclidean distances, the idea is even more applicable to higher dimensions.

If we are given space of locations or “states”, $\mathbf{R}^n$, it is our choice to pick an appropriate way to compute distances from coordinates:

Definition 1.3.12: three metrics

Suppose points P and Q in $\mathbf{R}^n$ have coordinates $(x_1, x_2, ..., x_n)$ and $(x'_1, x'_2, ..., x'_n)$ respectively.

1. The *Euclidean metric*, or the L^2-metric, is defined to be

$$d_2(P,Q) = \sqrt{\sum_{k=1}^{n} (x_k - x'_k)^2}$$

2. The *taxicab metric*, or the L^1-metric, is defined to be

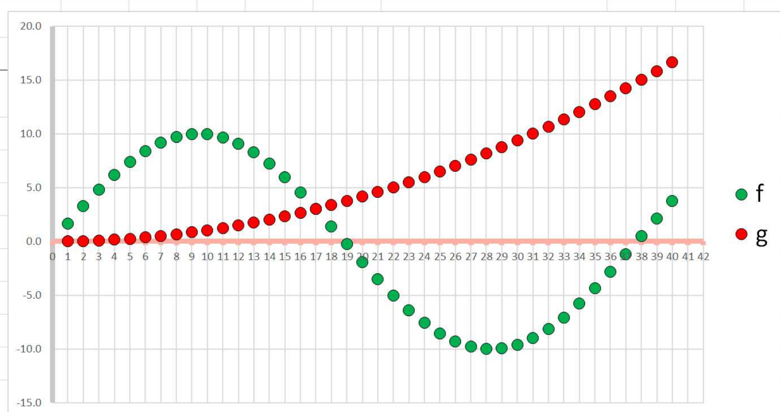
$$d_1(P,Q) = \sum_{k=1}^{n} |x_k - x'_k|$$

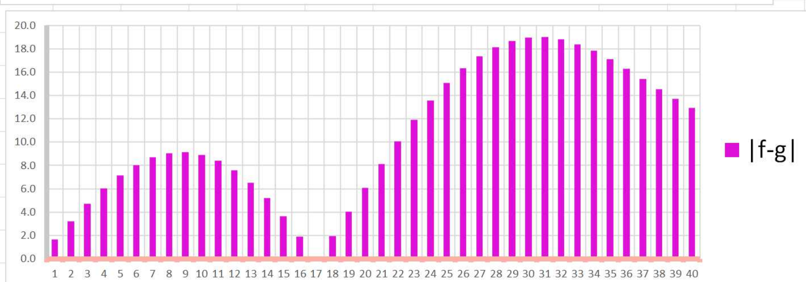
3. The *max metric*, or the L^∞-metric, is defined to be

$$d_\infty(P,Q) = \max_{k=1,\dots,n} |x_k - x'_k|$$

They are illustrated below ($n = 40$):

n=	40			L_1	L_2	L_∞
				TCM	EM	sup
	k	f	g	424.05	76.40	19.00
	1	1.66	0.01	1.65	2.72	
	2	3.27	0.04	3.23	10.43	
	3	4.79	0.09	4.70	22.09	
	4	6.18	0.17	6.02	36.20	
	5	7.40	0.26	7.14	51.00	
	6	8.41	0.38	8.04	64.64	
	7	9.19	0.51	8.68	75.41	
	8	9.72	0.67	9.05	81.95	
	9	9.97	0.84	9.13	83.38	
	10	9.95	1.04	8.91	79.43	
	11	9.66	1.26	8.40	70.51	
	12	9.09	1.50	7.59	57.65	
	13	8.28	1.76	6.52	42.46	
	14	7.23	2.04	5.19	26.93	
	15	5.98	2.34	3.64	13.26	
	16	4.57	2.67	1.91	3.63	
	17	3.03	3.01	0.02	0.00	
	18	1.41	3.38	1.96	3.86	





For the Euclidean metric, we compute for each row:

```
=ABS(RC[-2]-RC[-1])
```

We plot it, then apply this formula:

```
=SUM(R[1]C:R[40]C)
```

For the taxicab metric, we compute for each row:

```
=RC[-1]^2
```

We then apply this formula:

```
=SQRT(SUM(R[1]C:R[40]C))
```

The formula for max metric is simply:

```
=MAX(R[1]C[-2]:R[40]C[-2])
```

Exercise 1.3.13

Prove the Axioms of Metric Space for these formulas.

When we deal with the "physical space" ($n = 1, 2, 3$) as in the above theorems, the Euclidean metric is implied. For the "abstract spaces" ($n = 1, 2, ...$), the Euclidean metric is the default choice; however, there are many examples when the Euclidean geometry and, therefore, the Euclidean metric (aka the Distance Formula) don't apply.

Example 1.3.14: attributes

Let's consider the prices of wheat and sugar again. The space of prices is the same, $\mathbf{R}^2$. However, measuring the distance between two combinations of prices with the Euclidean metric leads to undesirable effects. For example, such a trivial step as changing the latter from "per ton" to "per kilogram" will change the geometry of the whole space. It is as if the space is stretched vertically. As a result, in particular, point P that used to be closer to point A than to B might now satisfy the opposite condition.

Furthermore, the two (or more) measurements or other attributes might have nothing to do with each other. In some obvious cases, they will even have different *units*. For example, we might compare two persons *built* based to the two main measurements: weight and height. Unfortunately, if we substitute such numbers into our formula, we will be adding pounds to feet!

Some of the concepts of geometry find their analogs in higher dimensions. For example, consider:

- A *circle* on the plane is defined to be the set of all points a given distance away from its center.
- A *sphere* in the space is defined to be the set of all points a given distance away from its center.

What about higher dimensions? The pattern is clear:

- A *hypersphere* in $\mathbf{R}^n$ is defined to be the set of all points a given distance away from its center.

In other words, each point P on the hypersphere satisfies:

$$d(P, Q) = R\,,$$

where Q is its center and R is its radius.

Example 1.3.15: Newton's Law of Gravity

According to the law, the force of gravity between two objects is
- proportional to either of their masses,
- inversely proportional to the square of the distance between their centers.

In other words, the force is given by the formula:

$$F = G\frac{mM}{r^2}\,,$$

where:
- F is the force between the objects;
- G is the gravitational constant;
- m is the mass of the first object;
- M is the mass of the second object;
- r is the distance between the centers of the mass of the two.

The dependence of F on m and M is very simple and, furthermore, we can assume that the masses of planets are remain the same. We are left with a function of one variable:

$$F(r) = G\frac{mM}{r^2}\,.$$

More precisely, r depends on the *location* P of the second object in the 3-dimensional space:

$$r = d(O, P)\,,$$

if, for simplicity, we assume that the first object is located at the origin. Note that this force is constant along any of the spheres centered at O.

Now, we can rewrite the law as a function of P:

$$\boxed{F(P) = \frac{GmM}{d(O,P)^2}}$$

Furthermore, if we suppose that the three spatial variables x, y, z are the coordinates of P, we can rewrite the law as a function of three variables:

$$F(x,y,z) = \frac{GmM}{d(O,P)^2} = \frac{GmM}{\left(\sqrt{x^2+y^2+z^2}\right)^2} = \frac{GmM}{x^2+y^2+z^2}\,.$$

If we ignore the third variable ($z = 0$), we can plot the graph of the resulting function of two variables:

But what about the *direction* of this force? This question is addressed in the next section.

Exercise 1.3.16

Visualize the function for the case of 3 dimensions.

1.4. Where vectors come from

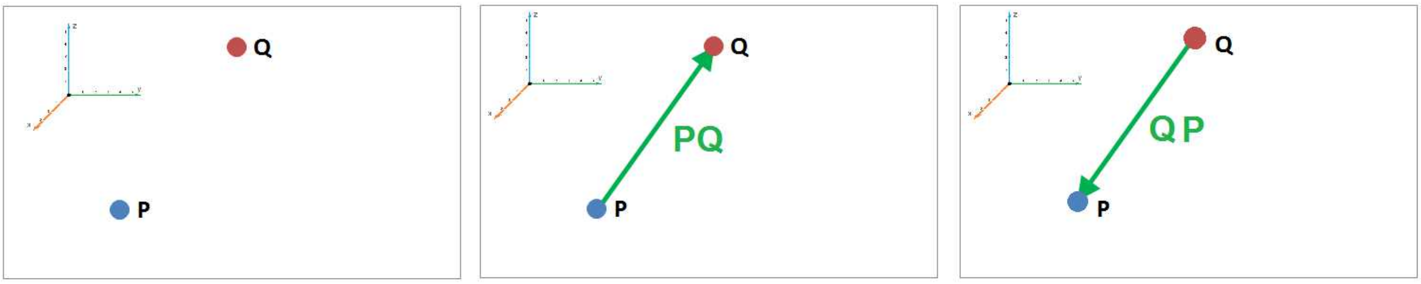
We introduced vectors in previously to properly handle the geometric issue of *directions* and angles between directions. However, vectors also appear frequently in our study of the natural world.

Definition 1.4.1: displacement

When the points in $\mathbf{R}^n$ are called locations or positions, the vectors are called *displacements*. In particular, if P and Q are two locations, then the vector PQ is the *displacement* from P to Q.

The idea applies to any space $\mathbf{R}^n$ but we will start with the physical space devoid of a Cartesian system.

From this point of view, a *vector* is a pair, PQ, of locations P and Q.



Warning!

"Vector" is not synonymous with "segment"; it's not even a set. The segment that you see is just a visualization.

We saw vectors in action previously, but the goal was limited to using vectors to understand directions and angles between them. Our interest here is the *algebraic operations* on vectors.

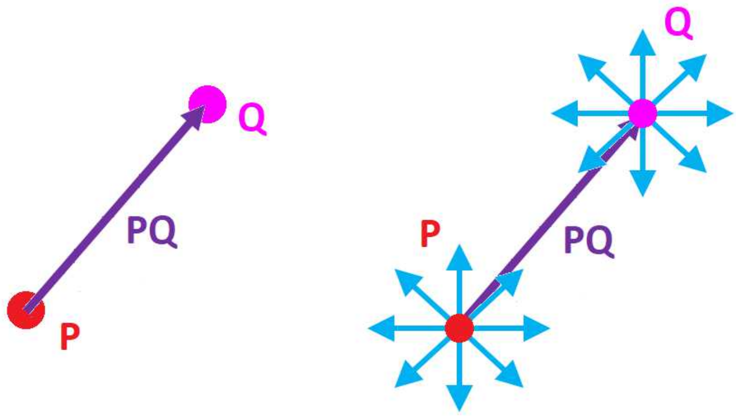
If a vector is an *ordered* pair, this means that $PQ \neq QP$. But is there a relation? The displacement from P to Q is the opposite to the displacement from Q to P:

$$\boxed{QP = -PQ}$$

The locations and displacements and, therefore, points and vectors are subject to algebraic operations that connect them:

$$\boxed{P + PQ = Q}$$

As you can see, we add a vector to a point that is its initial point and the result is its terminal point.



It follows:

$$PQ = Q - P\,.$$

As you can see, the vector is the difference of its terminal and its initial points. It follows that

$$QP = P - Q = -(Q - P) = -PQ\,.$$

We are back to the above formula.

Definition 1.4.2: affine space

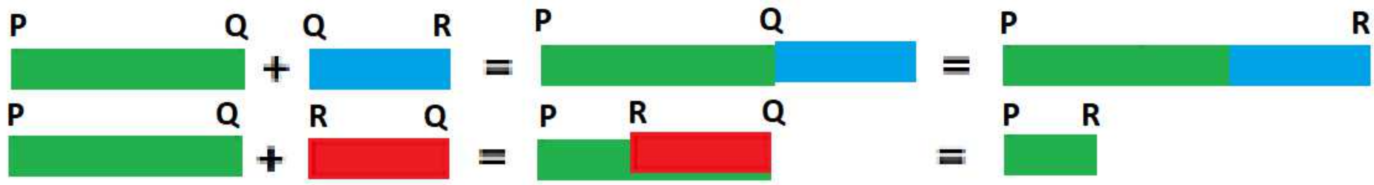
The *affine space* of the Euclidean space $\mathbf{R}^n$ is the set of ordered pairs PQ of points P and Q in this space. These pairs are called *vectors*.

We now review the algebra.

First, dimension 1.

Even though the algebra of vectors is the algebra of real numbers, we can still, even without a Cartesian system, think of the algebra of *directed segments*.

The *addition* of two vectors is executed by attaching the head of the second vector to the tail of the first, as illustrated below:

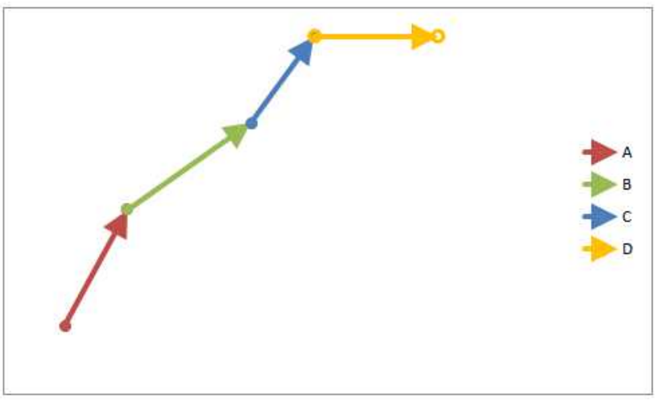


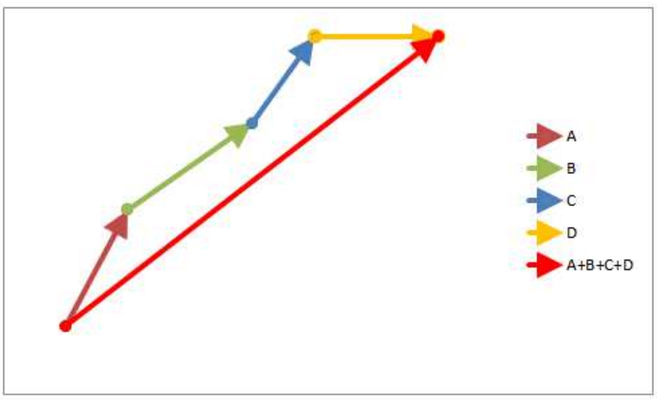
The negative number (red) is a segment directed backwards so that its tail is on its left.

Now dimension 2.

Example 1.4.3: consecutive displacements

We move point to point through the space:





This is how we can understand addition of vectors as displacements:

initial location	displacement	terminal location		
P	PQ	$P+PQ$	$=Q$	
Q	QR	$Q+QR$	$=R$	$=P+(PQ+QR)$
R	RS	$R+RS$	$=S$	$=P+(PQ+QR+RS)$
S	ST	$S+ST$	$=T$	$=P+(PQ+QR+RS+ST)$

The right column shows how adding vector to point can be seen as an alternative approach: We add vector to vector first.

For the general case of m steps, we have these two representations:

$$\begin{array}{llll}\sum_{k=0}^{m} X_kX_{k+1} & =X_0X_1+... & +X_{m-1}X_m & =X_0X_m\\ \sum_{k=0}^{m}(X_{k+1}-X_k) & =(X_1-X_0)+... & +(X_m-X_{m-1}) & =X_m-X_0\end{array}$$

Since moving from P to Q and then from Q to R amounts to moving from P to R, the construction is, again, a "head-to-tail" alignment of vectors:

$$PR=PQ+QR\,.$$

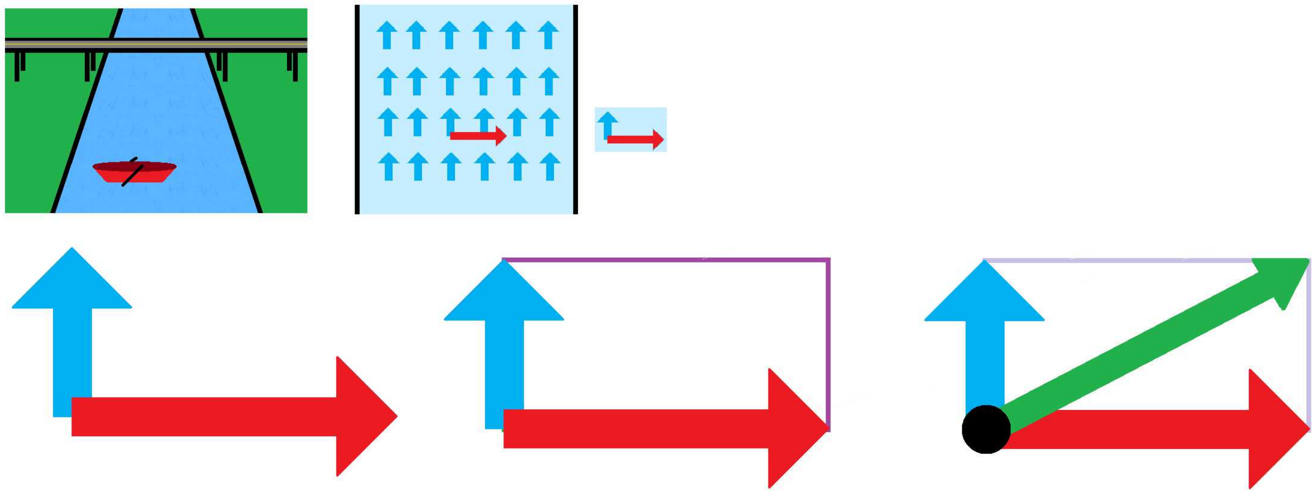
Warning!

We use the same symbol "+" as for addition of numbers.

However, in the physical world, there are other "metaphors" for vectors besides the displacements.

Example 1.4.4: velocity of stream

We look at the velocities of particles in a stream at each location. Then they may be combined with the speed of rowing of the boat:



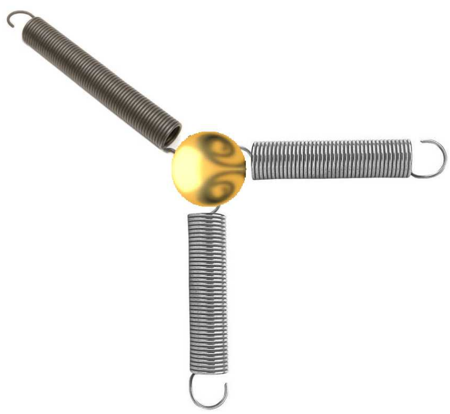
We are to add these two vectors at each location, but they, in contrast to the displacements, start at the *same* point!

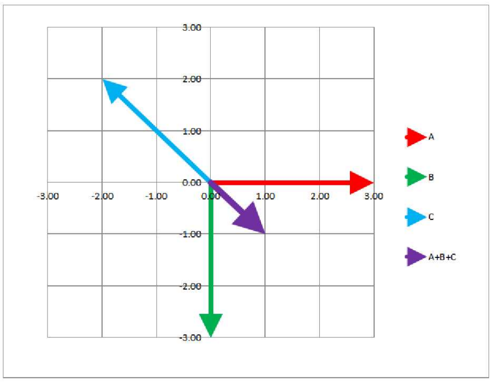
Exercise 1.4.5

With the velocities as shown, what is the best strategy to cross the canal?

Example 1.4.6: forces

Let's also look at *forces as vectors*. For example, springs attached to an object will pull it in their respective directions:





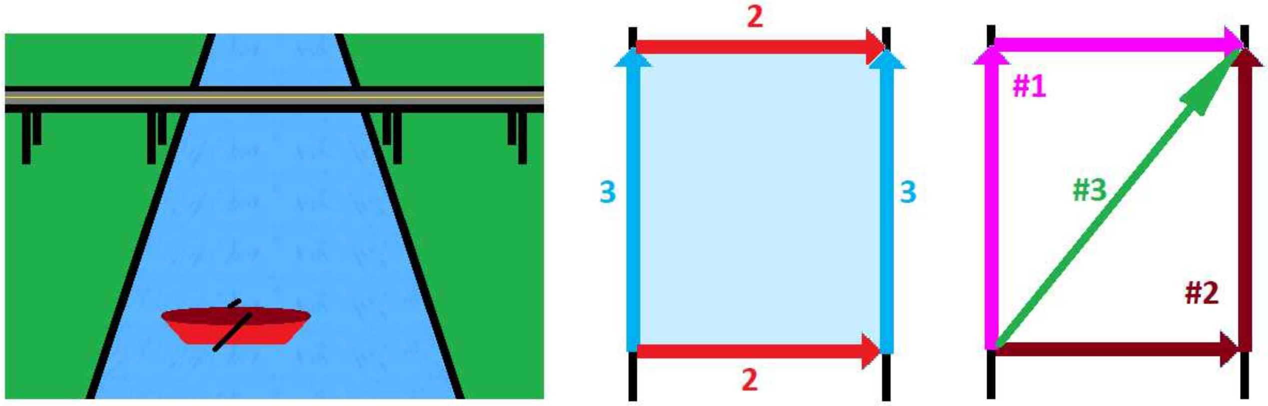
We add these vectors to find the combined force as if produced by a single spring. The forces are vectors that start at the same location.

Example 1.4.7: displacements

We can interpret the displacements, too, as vectors aligned to their starting points. Imagine we are crossing a river 3 miles wide and we know that the current (with no rowing) takes us 2 miles downstream. Three different ways this can happen:

1. a free-flow trip 3 miles north followed by a walk 2 miles east over a bridge; or
2. a walk 2 miles east over another bridge followed by a free-flow trip 3 miles north; but also
3. a rowing trip along the diagonal of a rectangle with one side going 3 miles north and another 2 miles east.

The three outcomes are the same:



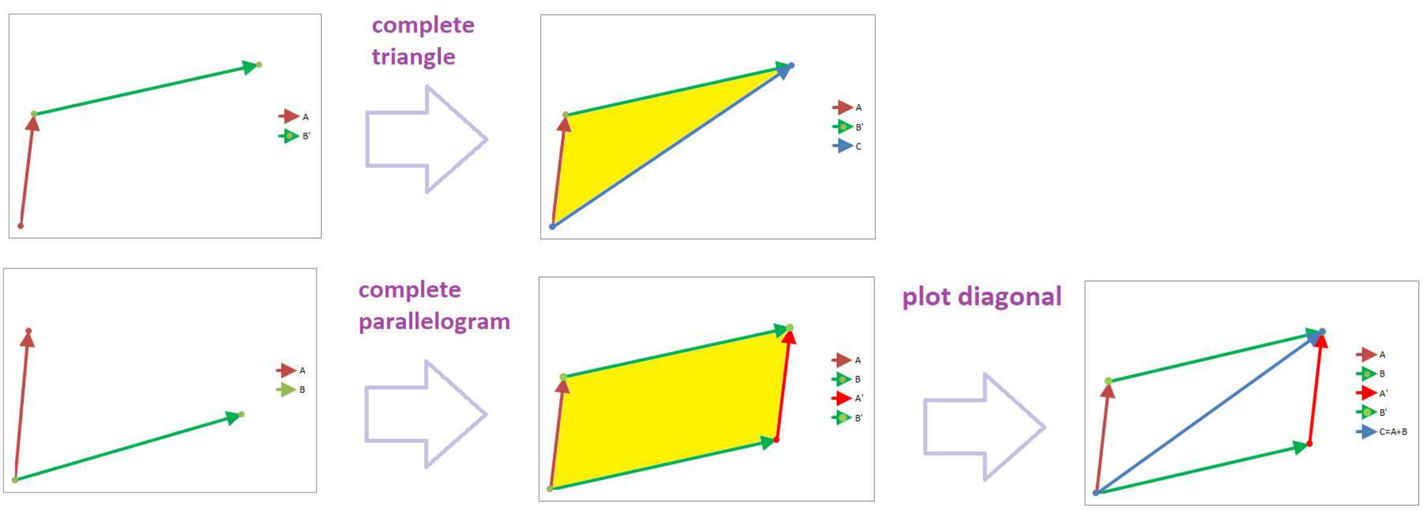
The sum of these two vectors is the same in any order:

3 miles north and 2 miles east

So, to add two vectors, we follow either

1. The head-to-tail: the triangle construction.
2. The tail-with-tail: the parallelogram construction.

They have to produce the same result! They do, as illustrated below:



1. For the former, we make a *copy* B' of B, attach it to the end of A, and then create a new vector with the initial point that of A and terminal point that of B'.
2. For the latter, we make a copy B' of B, attach it to the end of A, also make a copy A' of A, attach it to the end of B. Then the *sum* $A+B$ *of two vectors* A *and* B with the same initial point is the vector with the same initial point that is the diagonal of the parallelogram with sides A and B.

Exercise 1.4.8

Prove that the result is the same according to what we know from Euclidean geometry.

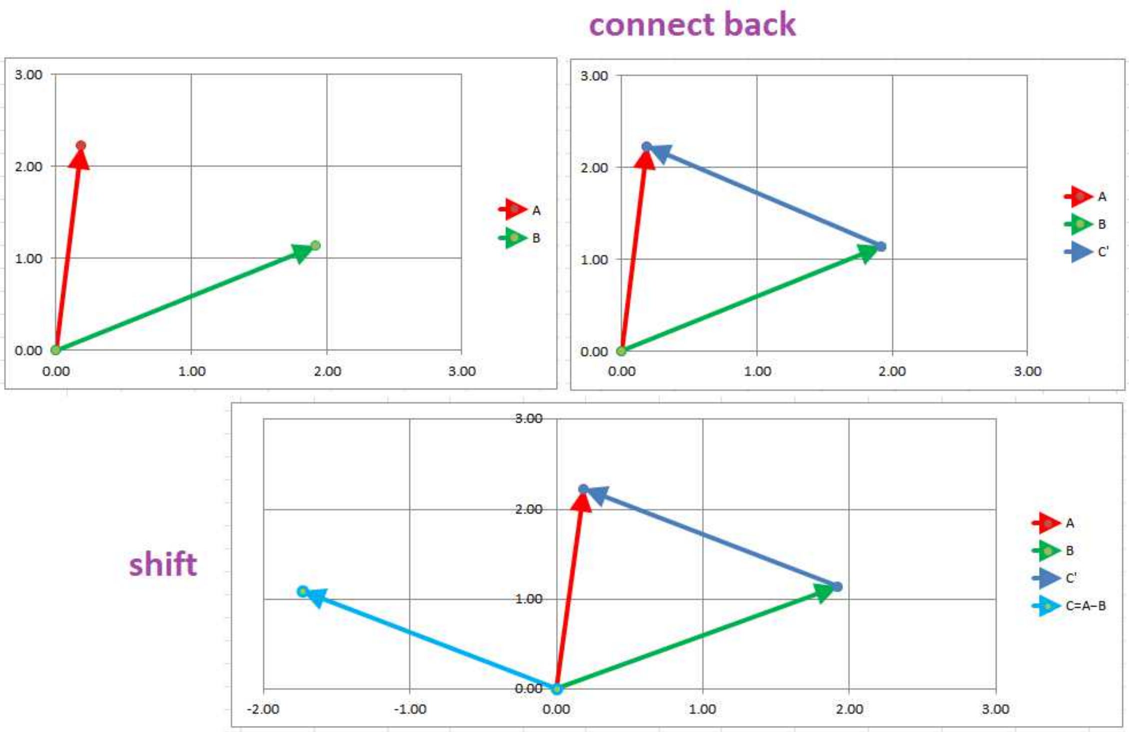
It is the same construction.

We think about vectors as line segments in a Euclidean space. As such, it has a direction and the length. It is possible to have the length to be 0; that's the *zero vector*. Its direction is undefined.

Once we know addition, *subtraction* is its inverse operation. Indeed, given vectors A and B, finding the vector C such that $B+C=A$ amounts to solving an equation, just as with numbers:

$$A-B=C \implies B+C=A\,.$$

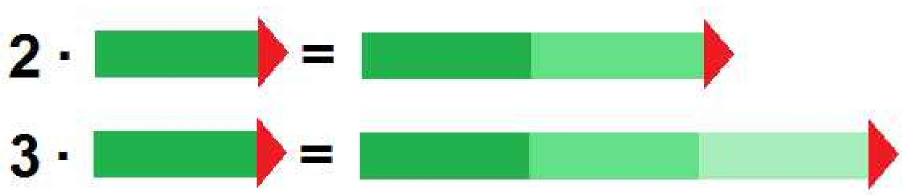
In other words, what do I add to B to get A? An examination reveals the answer:



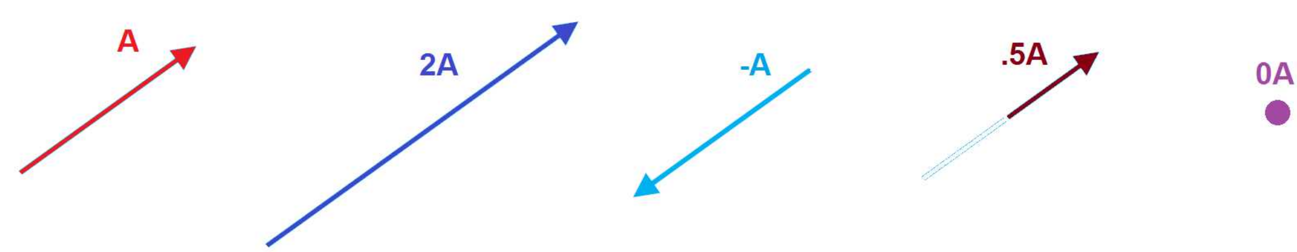
So, we construct the vector from the end of B to the end of A. One more step: make a copy of C with the same starting point as A.

If we want to go faster, we row twice as hard; the vector has to be stretched! Or, one can attach two springs in a consecutive manner to double the force, or cut any portion of the spring to reduce the force proportionally. A force might keep its direction but change its magnitude! It might also change the direction to the opposite.

There is then another algebraic operation on vectors. This is dimension 1:



This is dimension 2:



As you can see, every point has a special vector attached to it. For every point P, the *zero vector* is:

$$0 = PP\,.$$

We say that 0 serve as the *identity*.

Thus, the scalar product $c \cdot A$ of a vector A and a real number c is the vector with the same initial point as A, with the direction which is

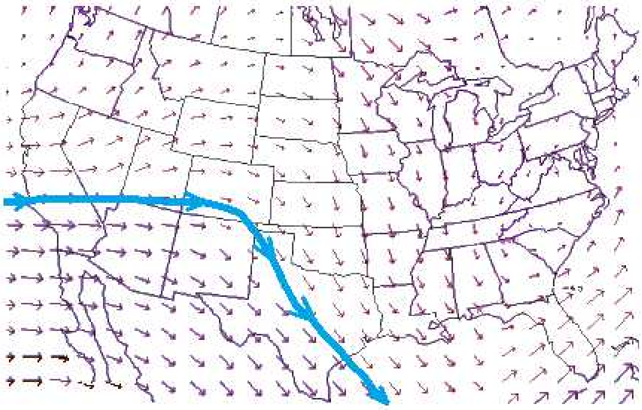
- same as that of A when $c > 0$,
- opposite to that of A when $c < 0$, and
- zero when $c = 0$.

Warning!

We use the same symbol "$\cdot$" as for multiplication of numbers.

Example 1.4.9: velocity of wind

Velocities appear as the wind speed at different locations:



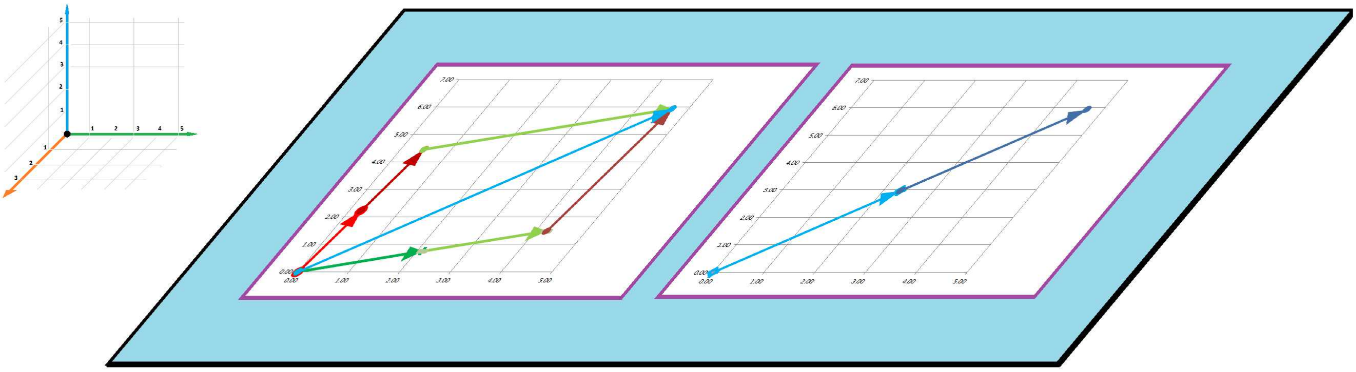
If the velocities are combined with the time increments, we can find the displacements of the particles of the air. We can also plot a whole trip of one such particle, just as in the beginning of the section. With the velocities denoted as follows $V_k,\ k=1,2,...,m$, respectively, takes the form:

$$\sum_{k=0}^{m} V_k \cdot \Delta t = X_m - X_0\,.$$

Exercise 1.4.10

Plot a few more paths.

What is the dimension of the space? As we know from Euclidean geometry, two lines and, therefore, two vectors, determine a *plane*. This is why we imagine that the operations, as we have defined them, are limited to a certain plane within a possibly higher-dimensional space:



Example 1.4.11: units

Out of caution, we should look at the units of the scalar. Yes, the force is a multiple of the acceleration:

$$F = ma\,.$$

However, these two have different units and, therefore, cannot be added together!

Also, the displacement is the time multiplied by the velocity:

$$\Delta X = \Delta t \cdot V\,.$$

But these two have different units and, therefore, cannot be added! They live in two different spaces.

We will continue, throughout the chapter, to use *capitalization* to help to tell vectors from numbers.

Warning!

To indicate vectors, many sources use:

- an arrow above the letter, $\vec{v}$, or
- the bold face, $\mathbf{v}$.

In this section, we introduced to $\mathbf{R}^n$, a space of points, a new entity – a vector. It is an ordered pair PQ of points, P and Q, linked back to points by this algebra:

$$\boxed{PQ = Q - P \quad \text{or} \quad Q = P + PQ}$$

The vectors can be added together and multiplied by a number according to the procedures described above.

The operations satisfy the following properties:

Theorem 1.4.12: Axioms of Affine Space

The points and the vectors in the affine space of $\mathbf{R}^n$ satisfy the following properties:

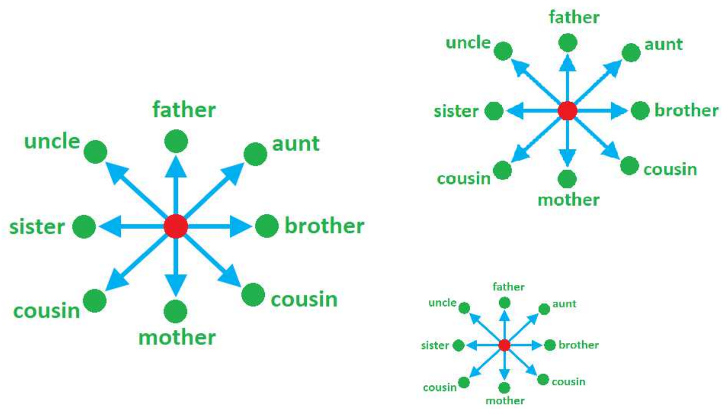
1. ***Identity****: For every point P, we have for some vector denoted as follows 0 the following:*
$$P + 0 = P\,.$$
2. ***Associativity****: For every point P and any vectors V and W starting at P, we have:*
$$(P + V) + W = P + (V + W)\,.$$
3. ***Free and transitive action****: For every point P and every point Q, there is a vector V such that*
$$P + V = Q\,.$$

Proof.

1. Choose:
$$0 = PP\,.$$
2. Compare:
$$(P + PQ) + QR = R \quad \text{and} \quad P + (PQ + QR) = P + PR = R\,.$$
3. Choose:
$$V = PQ\,.$$

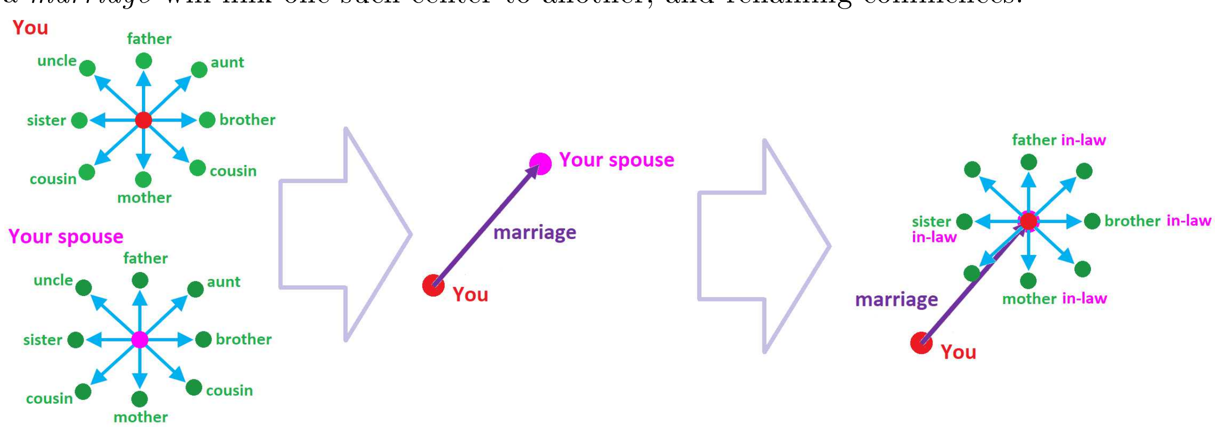
Example 1.4.13: family relations

Let's imagine that every point in the space stands for a person. Now, each person is linked by a vector to one's family:



Every person is the center of such a system of links. Also, potentially, each person is linked to any other person.

Now, a *marriage* will link one such center to another, and renaming commences:



The word "affine" means "in-law".

Let's simplify.

This is what we have in particular:

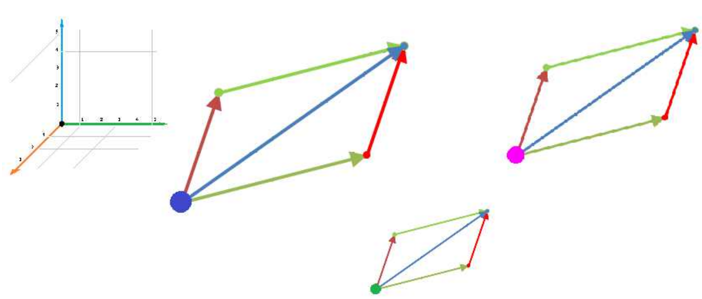
1. Two vectors with the same origin can be added together, producing another one with the same origin.
2. A vector can be multiplied by a number, producing another one with the same origin.

These operations are carried out according to the procedures described above. Given points P and Q and a real number k, we have for some points R and S:

$$\boxed{PQ + PR = PR \quad \text{and} \quad k \cdot PQ = PS}$$

Indeed, the result is another vector with the same initial point as the original(s)!

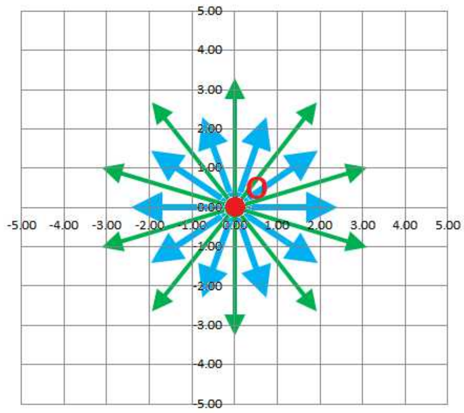
This algebra of vectors (including the scalar multiplication) can, therefore, be carried out separately at every location:



And this algebra is identical for every location! Therefore, a single initial point will be sufficient for our study of vector algebra.

But the choice of the initial point is crucial. This point will be assumed to be O, the origin, unless otherwise indicated.

This is what our collection of vectors looks like, for dimension 2:



Such a "space of vectors" is called a *vector space*. It is equipped with two operations, the vector addition:

$$\boxed{\text{vector} + \text{vector} = \text{vector}}$$

and the scalar multiplication:

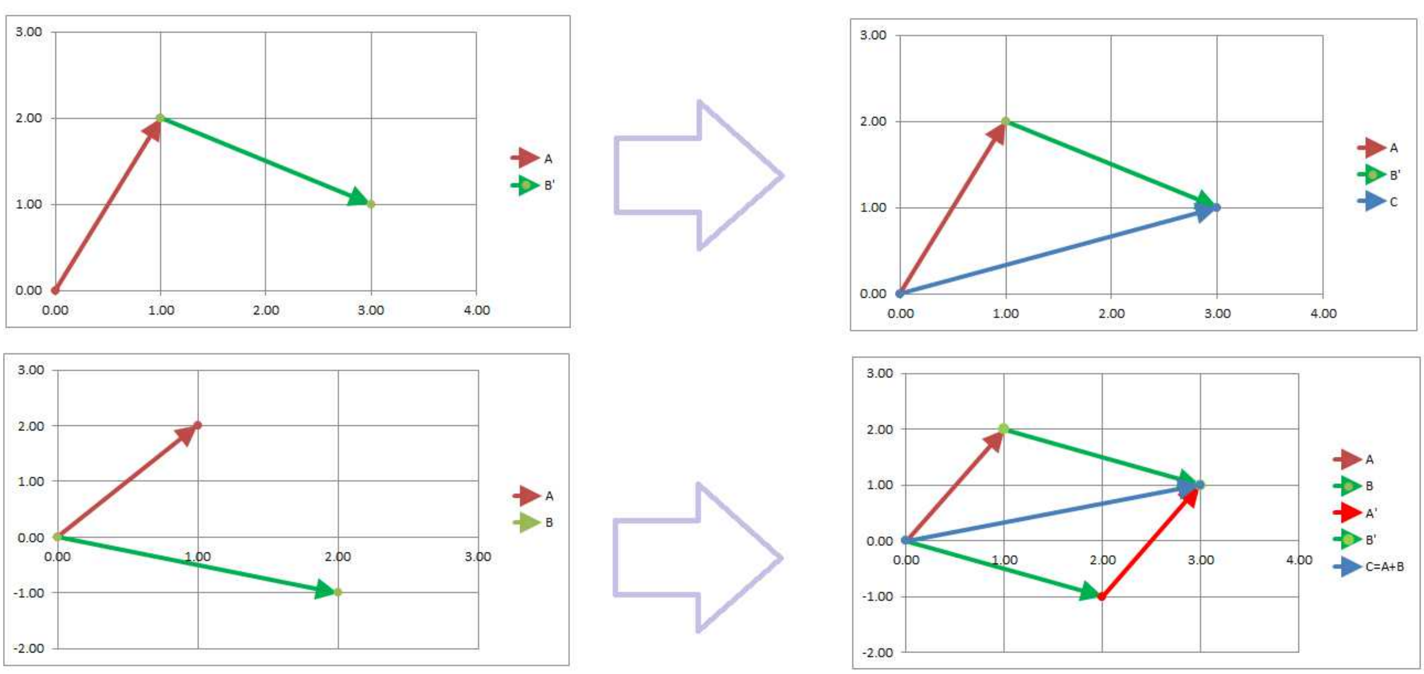
$$\boxed{\text{number} \cdot \text{vector} = \text{vector}}$$

Warning!

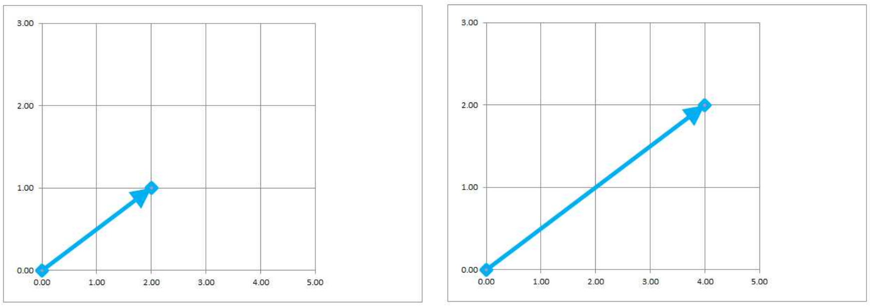
There are no points in a vector space.

Let's add a Cartesian system to our plane (with the origin already chosen).

We can watch what happens to the coordinate of the end-points of the vectors as we carry out our algebra. Sum implies coordinatewise addition:



Scalar product implies multiplication of the coordinates by the same number:



We are in $\mathbf{R}^2$ now.

1.5. *Vectors in* $\mathbf{R}^n$

We now understand vectors. Or at least we understand them in the lower-dimensional setting.

Now, we move to the next stage:

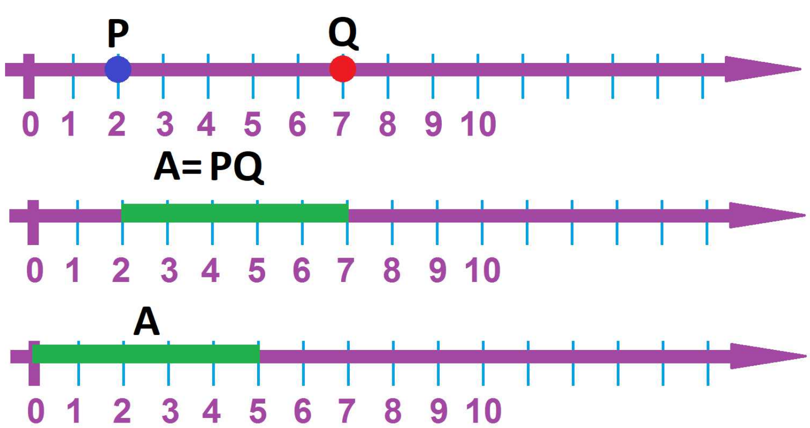
1. We add a Cartesian system to these spaces: line, plane, and space.
2. We also consider the abstract spaces of arbitrary dimensions, $\mathbf{R}^n$.

We will use the former to make sure that the approach to the latter makes sense.

A vector is still a pair PQ of points P and Q in $\mathbf{R}^n$. Now, either of these two points corresponds to a string of n numbers called its coordinates.

On the line $\mathbf{R}^1$, points are numbers and the vectors are simply differences of these numbers:

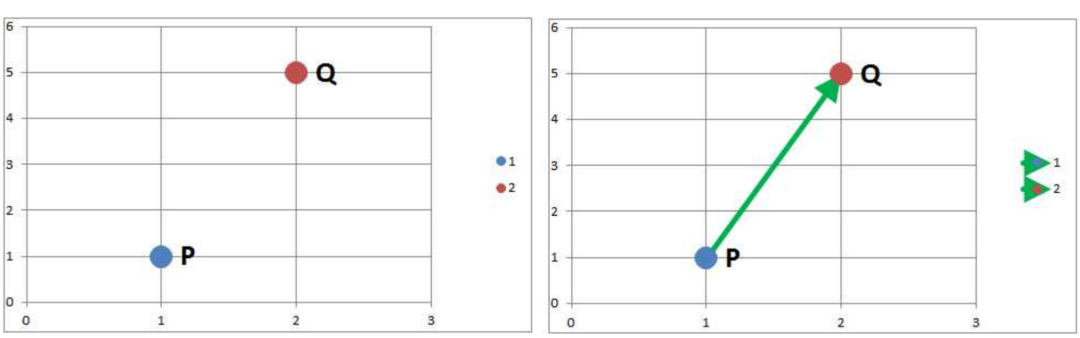
$$PQ = Q - P\,.$$



On the plane $\mathbf{R}^2$, we might have:

$$P = (1,2) \ \text{ and } \ Q = (2,5)\,.$$

How can we express vector PQ in terms of these four numbers?



We look at the *change* from P to Q:

- The change with respect to x, which is $2-1=1$.
- The change with respect to y, which is $5-2=3$.

We combine these into a new *pair of numbers* (with triangular brackets to distinguish these from points):

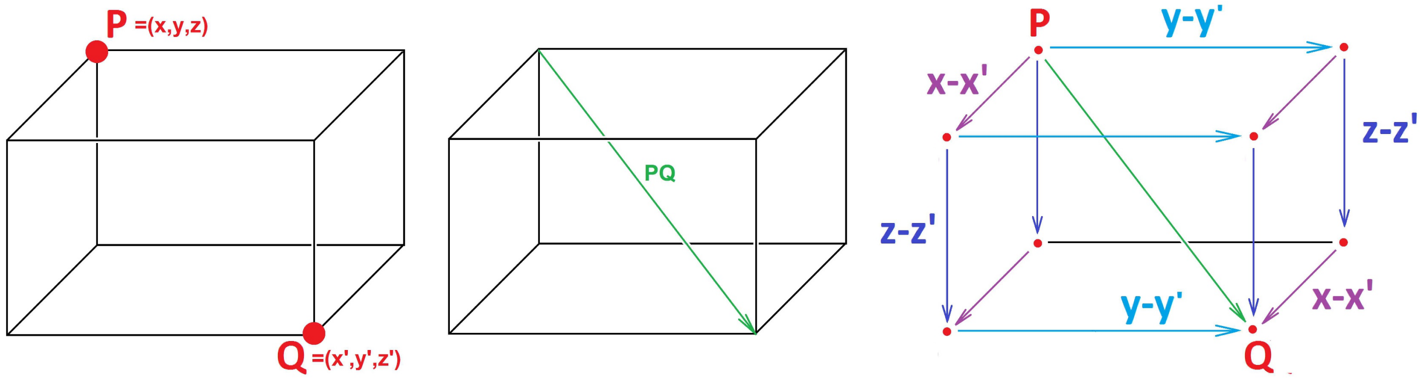
$$PQ = <1,3> \ .$$

Technically, however, we have to mention the initial point $P=(1,2)$ of the vector.

Now in $\mathbf{R}^3$, we might have:

$$P=(x,y,z) \ \text{ and } \ Q=(x',y',z')\,.$$

How can we express vector PQ in terms of these six numbers?



There are three changes (differences) along the three axes, i.e., a *triple*:

$$PQ = <x'-x, y'-y, z'-z> \ .$$

Definition 1.5.1: vector and its components

A *vector* PQ in $\mathbf{R}^n$ with its initial point

$$P=(x_1,x_3,...,x_n)$$

and its terminal point

$$Q=(x_1',x_2',...,x_n')$$

is given by the string of n numbers called the *components* of the vector:

$$x_1'-x_1,\ x_2'-x_3,\ ...,\ x_n'-x_n\,.$$

The definition matches the one that relies on directed segments.

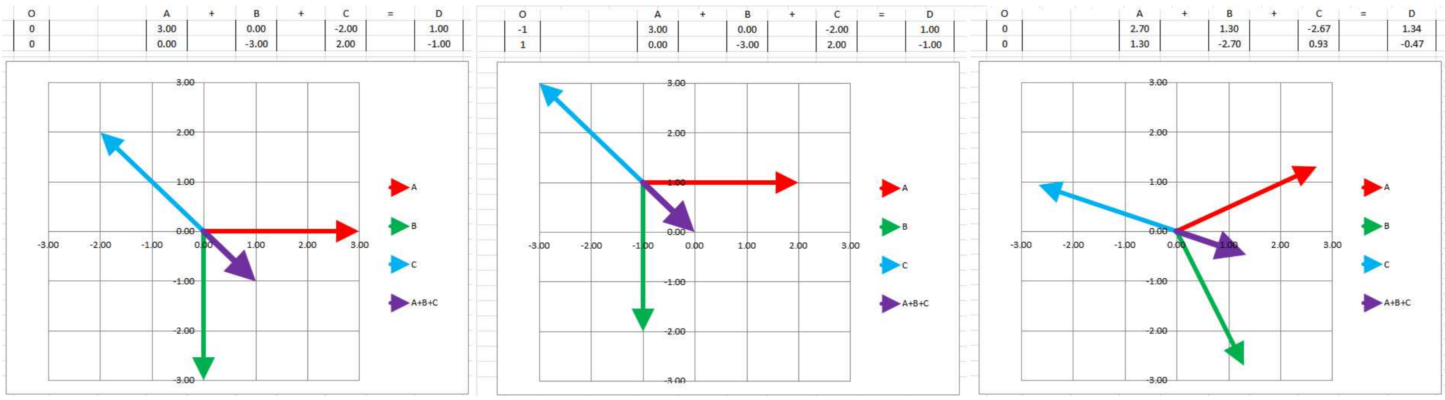
A vector may emerge from its initial and terminal points or independently. In either case, we assemble the components according to the following notation.

Row and column vectors

$$\left[\begin{array}{c} a_1 \\ a_2 \\ ... \\ a_n \end{array}\right] = <a_1,a_2,...,a_n>$$

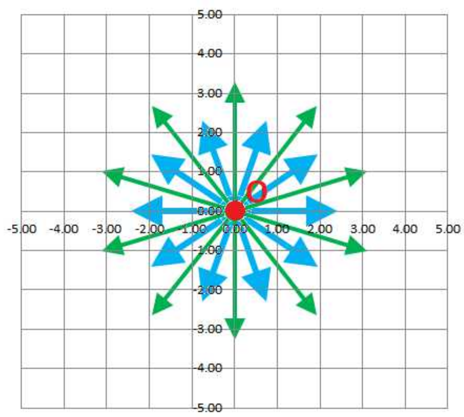
The former is preferred; the latter is its abbreviation.

Once again we can only carry out vector addition on vectors with the same initial point. What happens if we change the initial point while leaving the components of the vectors intact? Not only is each vector "copied" but so are the results of all algebraic operations. They are the same, just shifted to a new location:



We can even rotate the coordinate system (right).

It is then sufficient to provide results for the vectors that start at the *origin* O only! Only these are allowed:



In that case, *the components of a vectors are simply the coordinates of its end*:

$$\begin{array}{rl} P & = (x_1, \quad x_3, \quad ..., \quad x_n) \quad \Longrightarrow \\ OP & =< x_1, \quad x_2, \quad ..., \quad x_n > \end{array}$$

Warning!

The difference between points and vectors lies in the algebraic operations to which they are subject.

Next we consider the familiar algebraic operations but this time the vectors are represented by their components.

We carry out operations *componentwise*.

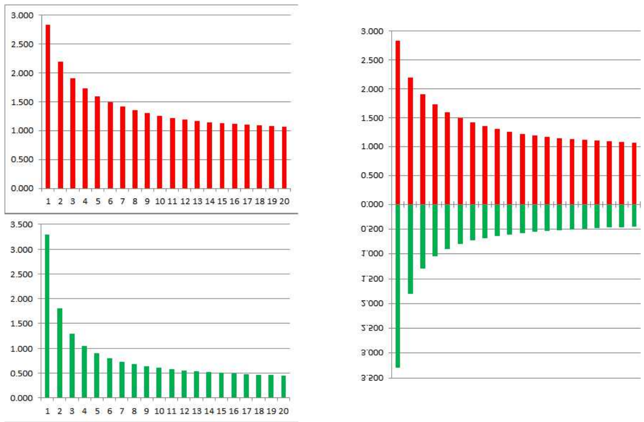
We demonstrate these operations for dimension $n = 3$ and for both the row and the column styles of notation. The vector addition is defined by:

$$\begin{array}{rlllll} A & =< & x, & y, & z & > \\ + & & & & & \\ B & =< & u, & v, & w & > \\ \hline A+B & =< & x+u, & y+v, & z+w & > \end{array} \qquad \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} x+u \\ y+v \\ z+w \end{bmatrix}$$

This is how we progress to the definition that doesn't involve points:

$$\boxed{\begin{array}{ll} P=(x_1,...,x_n) & OP \\ Q=(y_1,...,y_n) & OQ \\ \hline R=(x_1+y_1,...,x_n+y_n) & OP+OQ=OR \end{array}} \quad \to \quad \boxed{\begin{array}{ll} U & =<x_1,...,x_n> \\ V & =<y_1,...,y_n> \\ \hline U+V & =<x_1+y_1,...,x_n+y_n> \end{array}}$$

A visualization of this operation in $\mathbf{R}^{20}$ is below:



The scalar multiplication is defined by:

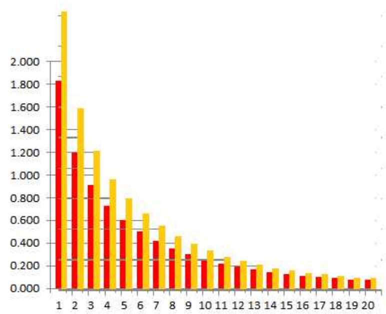
$$\begin{array}{lccccc} A & =< & x, & y, & z & > \\ \times & & & & & \\ k & & & & & \\ \hline kA & =< & kx, & ky, & kz & > \end{array} \qquad k \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ kz \end{bmatrix}$$

In either case, the components are aligned. Even though both seem equally convenient, the former will be seen as an abbreviation of the latter.

This is how we progress to the definition that doesn't involve points:

$$\boxed{\begin{array}{ll} k \text{ real} & \\ P=(x_1,...,x_n) & OP \\ \hline R=(kx_1,...,kx_n) & k\cdot OP=OR \end{array}} \quad \to \quad \boxed{\begin{array}{ll} k \text{ real} & \\ U & =<x_1,...,x_n> \\ \hline k\cdot U & =<kx_1,...,kx_n> \end{array}}$$

A visualization of this operation in $\mathbf{R}^{20}$ is below ($k=1.3$):



The scalar k is also known as the *constant multiple*.

Example 1.5.2: investment portfolios

If there are $10,000$ stocks on the stock market, every investment portfolio can be seen as a $10,000$-dimensional vector.

Then, *merging* two or more portfolios will *add* their vectors:

		$	$	$
		A	B	A+B
1	AGTK	20	3	23
2	AKAM	10	0	10
3	BCOR	5	2	7
4	BIDU	11	22	33
5	BRNW	12	10	22
6	CARB	15	0	15
7	CCIH	0	0	0
8	CCOI	0	6	6
9	JRJC	1	2	0
10	WIFI	23	2	25
11	...	...	...	...

We use the formula:

```
=RC[-2]+RC[-1]
```

Second, *doubling* or tripling a portfolio while preserving the proportion (or weight) of each stock will *scalar multiply* its vector:

		$	$
		A	2A
1	AGTK	20	40
2	AKAM	10	20
3	BCOR	5	10
4	BIDU	11	22
5	BRNW	12	24
6	CARB	15	30
7	CCIH	0	0
8	CCOI	0	0
9	JRJC	1	2
10	WIFI	23	46
11	...	...	...

We use the formula:

```
=2*RC[-1]
```

Even *non-homogeneous* holdings are subject to these operations:

$$< 10000 \text{ tons of wheat }, 20000 \text{ barrels of oil }, ... >,$$

or

$$< \$100000,\ ¥1000000,\ ...\ > \ .$$

Definition 1.5.3: sum of vectors

For two vectors in $\mathbf{R}^n$, their *sum* is defined to be the vector acquired by their componentwise addition:

$$\begin{array}{lll} & A & =< a_1, ..., a_n > \\ & B & =< b_1, ..., b_n > \\ \Longrightarrow & A + B & =< a_1 + b_1, ..., a_n + b_n > \end{array}$$

Definition 1.5.4: scalar product

For a number k and a vector in $\mathbf{R}^n$, their *scalar product* is defined to be the vector acquired by their componentwise multiplication:

$$\boxed{\begin{array}{rll} & k & \text{real} \\ & A & =< a_1, ..., a_n > \\ \Longrightarrow & kA & =< ka_1, ..., ka_n > \end{array}}$$

We thus have the algebra of vectors for a space of any dimension!

These operations can be proven to satisfy the same properties as the vectors in $\mathbf{R}^3$ (next section). The proof is straight-forward and relies on the corresponding property of real numbers. For example, to prove the commutativity of vector addition, we use the commutativity of addition of numbers as follows:

$$A + B = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} x+u \\ y+v \\ z+w \end{bmatrix} = \begin{bmatrix} u+x \\ v+y \\ w+z \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} + \begin{bmatrix} x \\ y \\ z \end{bmatrix} = B + A\,.$$

As a result, we are able to treat vectors as if they were numbers.

Furthermore, the *special* vectors deserve special attention:

Definition 1.5.5: zero vector

The *zero vector* in $\mathbf{R}^n$ has only zero components:

$$\boxed{0 =< 0, ..., 0 >}$$

Definition 1.5.6: the negative of vector

The *negative vector* of a vector A in $\mathbf{R}^n$ has its components the negatives of those of A:

$$\boxed{-A =< -a_1, ..., -a_n >}$$

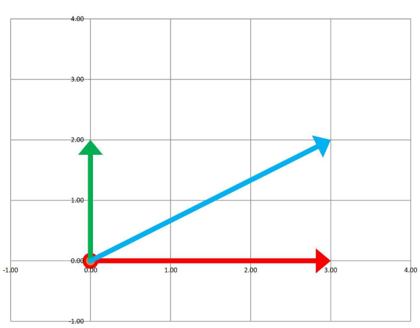
Exercise 1.5.7

Prove the eight axioms of vector spaces for the algebraic operations defined this way for (a) $\mathbf{R}^3$, (b) $\mathbf{R}^n$.

Let's explain the reason for the word "component".

A vector A is *decomposed* into the sum of other vectors. We chose those vectors to be special: Each is aligned with one of the axes. For example, we decompose:

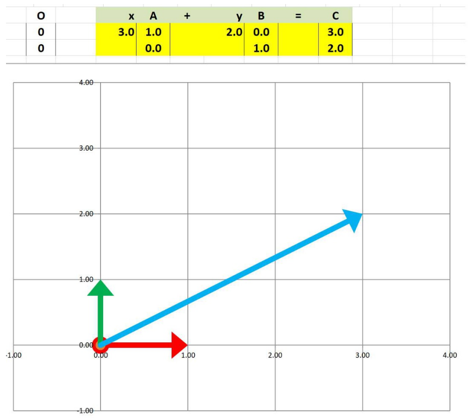
$$A =< 3, 2 > = < 3, 0 > + < 0, 2 > .$$



Then the two vectors are called the *component vectors* of A. We take this analysis one step further with scalar multiplication:

$$A =< 3,2 > =< 3,0 > + < 0,2 >= 3< 1,0 > + 2< 0,1 > .$$

This is a *decomposition* of A:



Similarly, *any* vector can be represented in such a way:

$$< a,b >= a < 1,0 > +b < 0,1 > \ .$$

We use the following notation for these special vectors in $\mathbf{R}^2$:

Basis vectors in R^2

$$i =< 1,0 >, \ j =< 0,1 >$$

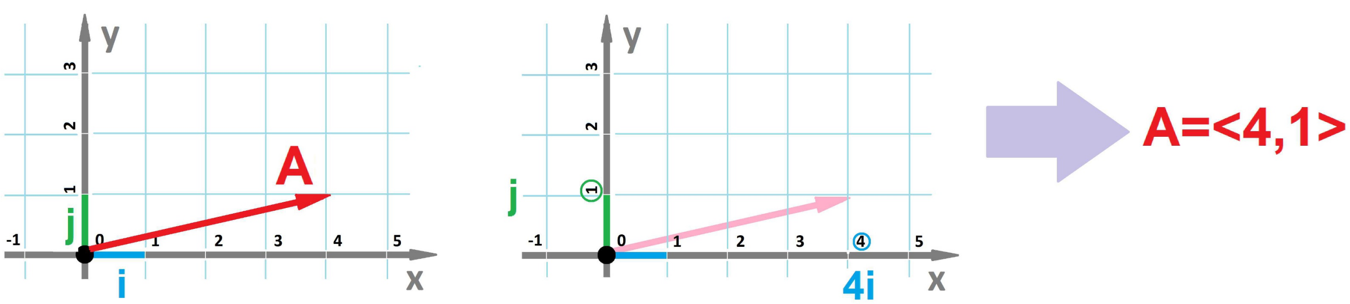
Then, any vector is as a *linear combination* of these two:

$$\boxed{< a,b >= ai + bj}$$

Example 1.5.8: decomposition

Below we present a new point of view. Before, we'd consider a point and its coordinates:

$$P = (4,1) \, .$$



Now we write a vector and its components:

$$4i + j =< 4,1 > \ .$$

Thus, representing a vector in terms of its components is just a way (a single way, in fact) to represent it in terms of a pair of specified unit vectors aligned with the axis.

We, furthermore, use the following notation for such vectors in $\mathbf{R}^3$:

Basis vectors in R^3

$$i =< 1,0,0 >,\ j =< 0,1,0 >,\ k =< 0,0,1 >$$

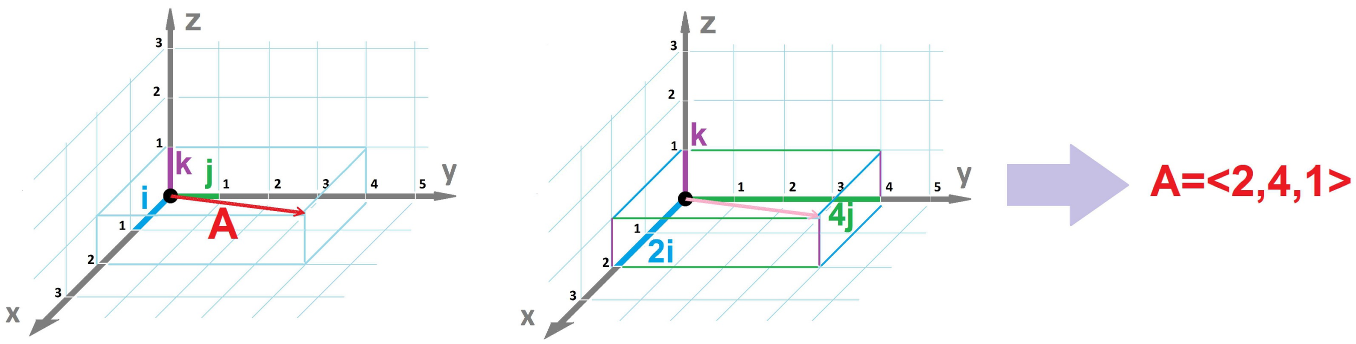
For every vector, we have the following representation:

$$< a,b,c >= ai + bj + ck$$

Example 1.5.9: decomposition

Now we write a vector and its components:

$$2i + 4j + k =< 2,4,1 > .$$



Definition 1.5.10: basis vectors

The *basis vectors* in $\mathbf{R}^n$ are defined and denoted as follows

$$\begin{aligned} &e_1 =< 1,0,0,...,0 > \\ &e_2 =< 0,1,0,...,0 > \\ &... \\ &e_n =< 0,0,0,...,1 > \end{aligned}$$

Together they form a *basis* of $\mathbf{R}^n$.

Then, any vector is as a *linear combination* of these n vectors:

$$< x_1, x_2, ..., x_n >= x_1e_1 + x_2e_2 + ... + x_ne_n$$

We have come to a new understanding:

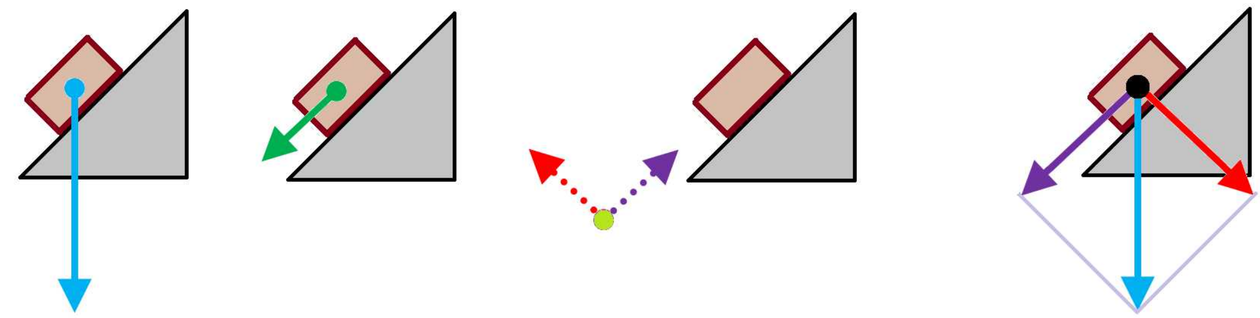
- old: Cartesian system = the axes
- new: Cartesian system = the origin and the basis vectors

Of course, we can choose a different Cartesian system by choosing a new set of basis vectors. The choice of the basis vectors is dictated by the problem to be solved.

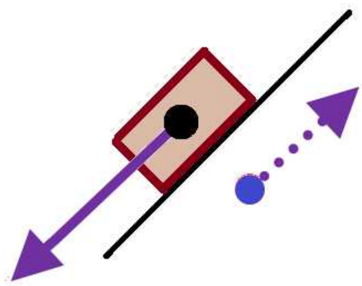
Example 1.5.11: compound motion

Suppose we are to study the motion of an object sliding down a slope.

Even though gravity is pulling it vertically down, the motion is restricted to the surface of the slope. It is then beneficial to choose the first basis vector i to be parallel to the surface and the second j perpendicular:



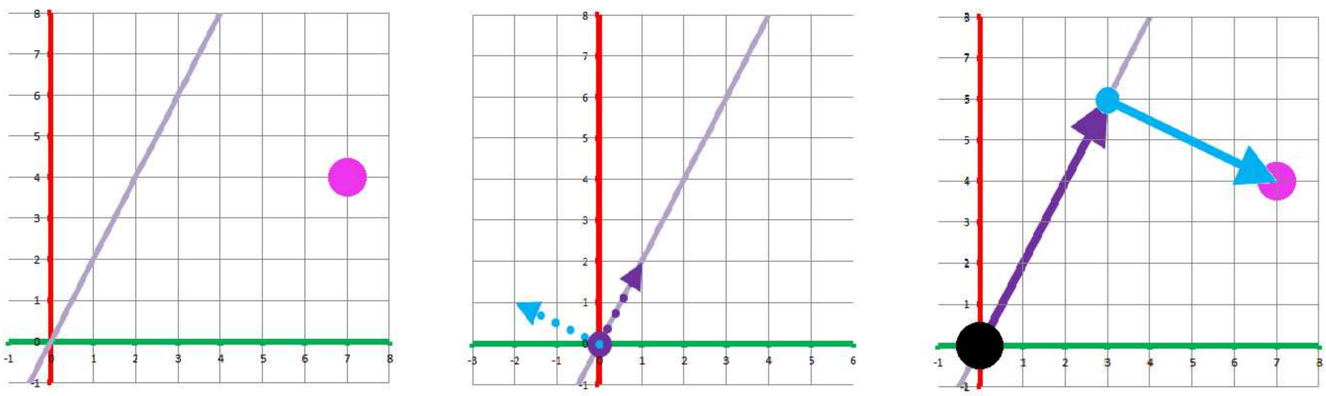
The gravity force is then decomposed into the sum of two vectors. It is the first one that affects the object and is to be analyzed in order to find the acceleration (Newton's Second Law). The second is perpendicular and is canceled by the resistance of the surface (Newton's Third Law). The problem becomes 1-dimensional:



Example 1.5.12: investing

Even when we deal with the abstract spaces $\mathbf{R}^n$, such decompositions may be useful.

For example, an investment advice might be to hold the proportion of stocks and bonds 1-to-2. We plot each possible portfolio as a point on the xy-plane, where x is the amount of stocks and y is the amount of bonds in it. Then the "ideal" portfolios lie on the line $y = 2x$. Furthermore, we would like to evaluate how well portfolios follow this advice. We choose the first basis vector to be $i =< 2, 1 >$ and the second perpendicular to it, $j =< -1, 2 >$:



Then the first coordinate – with respect to this new coordinate system – of your portfolio reflects how far you have followed the advice, and the second how much you've deviated from it. Now we just need to learn how to compute distances and angles in such a space.

1.6. *Algebra of vectors*

We will look for similarities with the algebra of numbers: the *laws of algebra*.

For example, we freely use the following shortcut when we deal with numbers:

$$x \cdot (y+z) = (x \cdot y) + (x \cdot z)\,.$$

Is there a similar rule for vectors? Yes, in a sense.

The complexity and the number of these laws will be higher because the participants are of two different types: numbers and vectors. They are also intermixed. For example, we may write an analog of the above formulas as follows:

$$x \cdot (Y+Z) = (x \cdot Y) + (x \cdot Z)\,.$$

Here x is still a number, but Y and Z are vectors. The formula can, however, be easily verified in specific situations. The left-hand side:

$$2 \cdot (<3,4> + <5,6>) = 2 \cdot (<3+5, 4+6> = 2 \cdot <8,10> = <16,20> \ .$$

The right-hand side:

$$2 \cdot <3,4> + 2 \cdot <5,6> = <2 \cdot 3, 2 \cdot 4> + <2 \cdot 5, 2 \cdot 6> = <6,8> + <10,12> = <6+10, 8+12> = <16,20> \ .$$

We will first explore these rules and short-cuts as they appear independently from componentwise representations of vectors.

In dimension 1, we deal with the algebra of directed segments. As these segments now all start at 0, this *is* the algebra of real numbers. Nonetheless, we keep the two types apart, with an eye on the higher dimensions.

The following simple idea connects vector addition to scalar multiplication:

$$A + A = 2A\,.$$

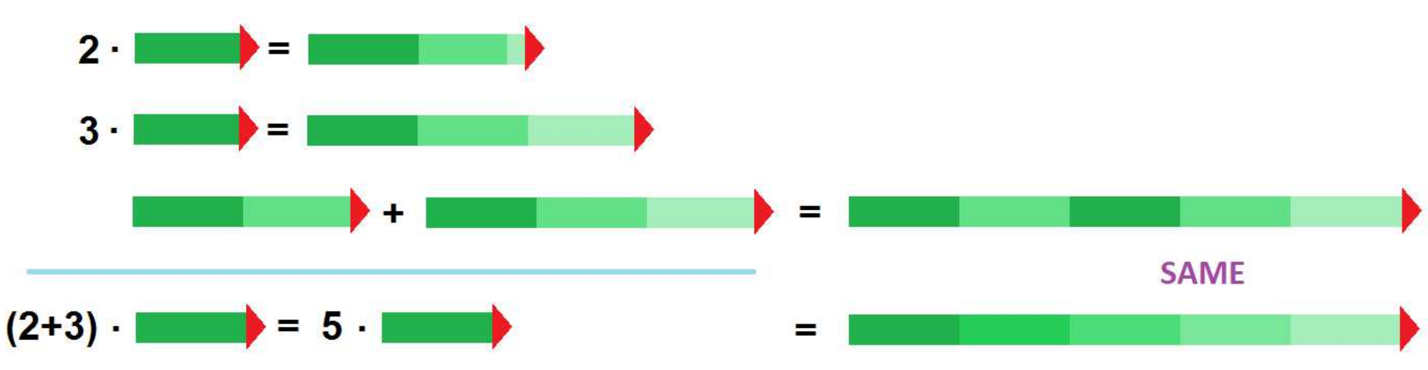
Its generalization is the *First Distributivity Property of Vector Algebra*:

$$\boxed{aA + bA = (a+b)A}$$

It's just factoring:

▶ We factor a *vector* out.

For example, let's add a double to a triple:



The result is the same if we quintuple the original vector.

In other words, we distribute scalar multiplication over addition of real numbers.

If we are to be precise, the symbol "+" stands for two different things above:

- $aA + bA$, addition of vectors
- $a + b$, addition of numbers

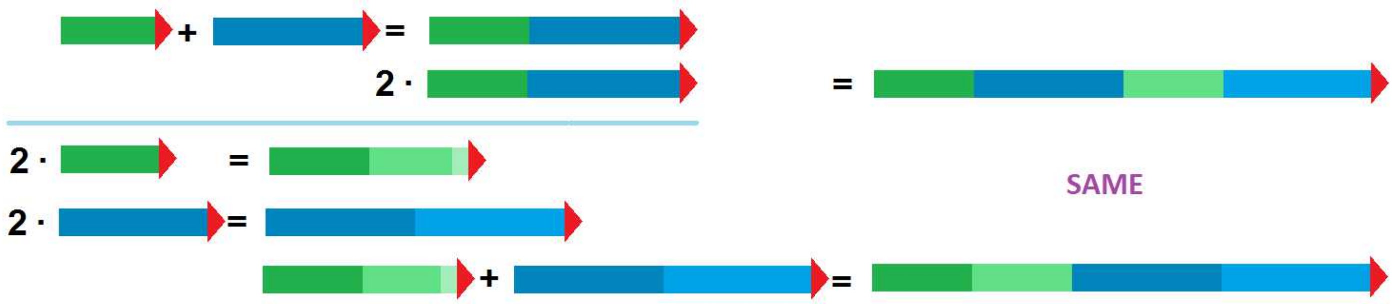
Next, we can also distribute multiplication of real numbers over addition of vectors. The *Second Distributivity Property of Vector Algebra* is:

$$\boxed{aA + aB = a(A + B)}$$

It's just factoring again:

- ▶ We factor a *number* out.

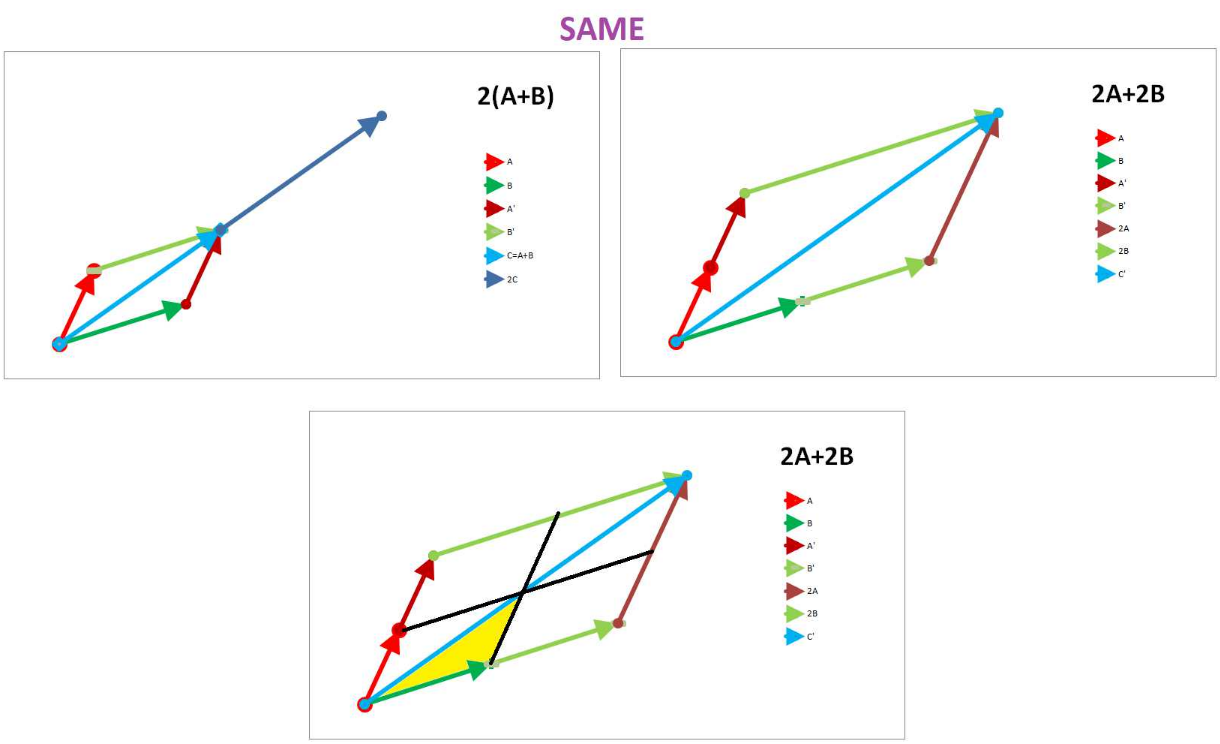
For example, let's double the sum:



The result is the same if we add the doubles.

Dimension 2.

Below, we add two vectors and then stretch the result (left) and we stretch two vectors and then add them (right):



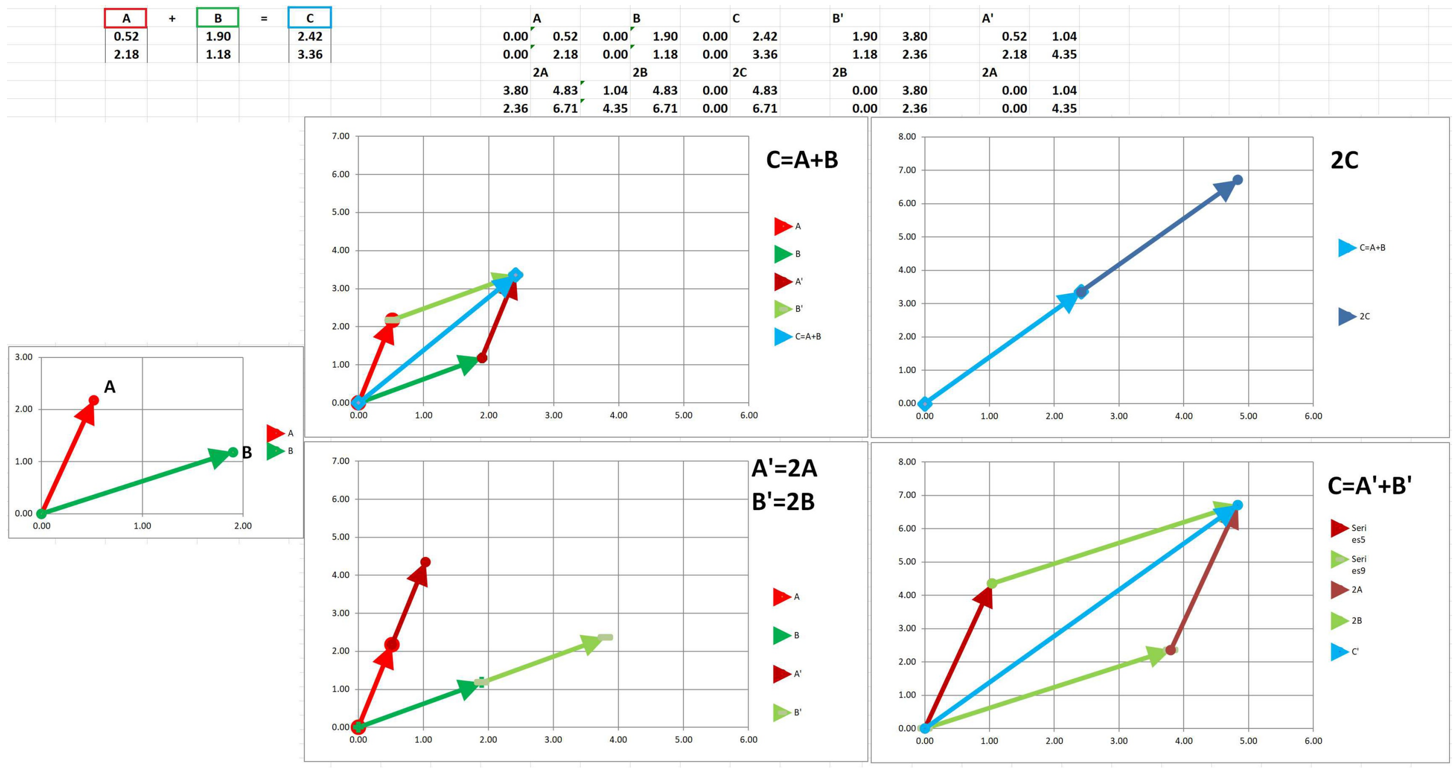
The result is the same.

Exercise 1.6.1

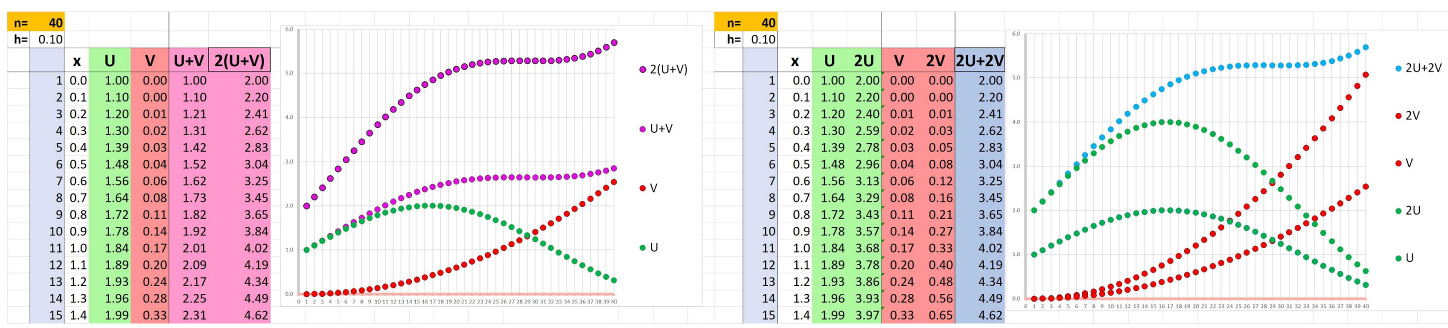
Explain why the results are the same. Hint: Similar triangles.

This algebraic rule, and others still to come, has been justified following the familiar algebra and geometry of the "physical space" $\mathbf{R}^2$ and $\mathbf{R}^3$. However, they also serve as a *starting point* for further development of linear algebra. In the last section, we defined the algebra of the abstract space $\mathbf{R}^n$ and now demonstrate that these "axioms" are still satisfied.

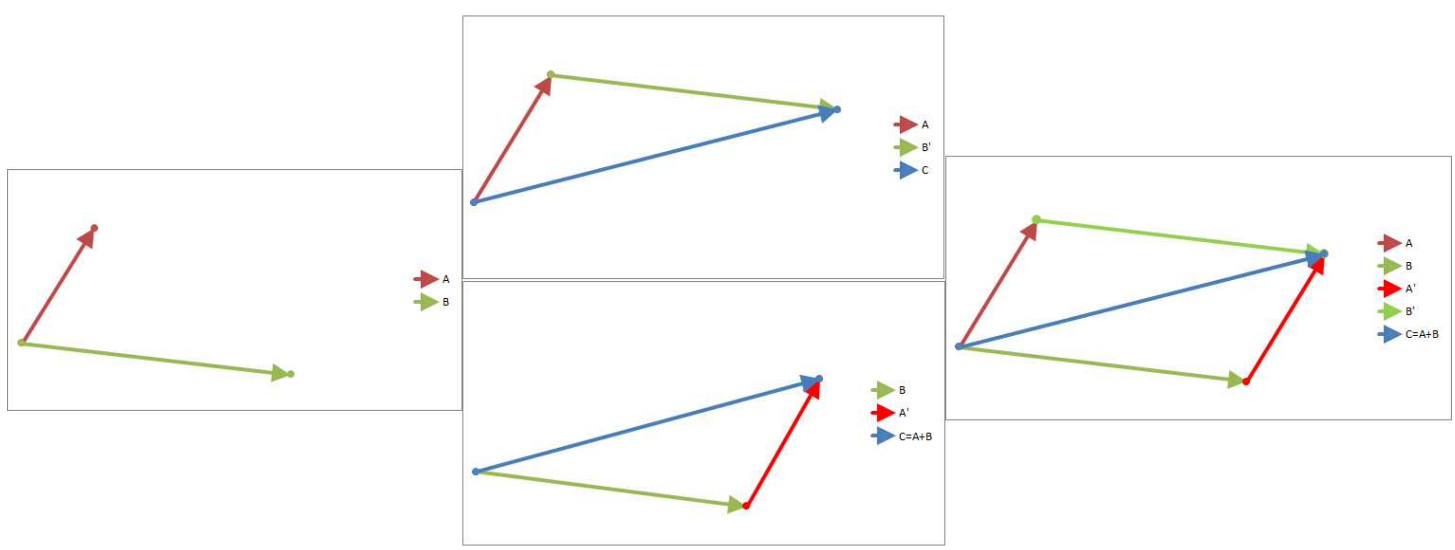
Below is the same formula for dimension 2, presented along with the components of the vectors:



The formula is illustrated for dimension 40:



Recall that to find $A+B$, we make a *copy* B' of B, attach it to the end of A, and then create a new vector with the initial point that of A and terminal point that of B'. Now, to find $B+A$, we make a *copy* A' of A, attach it to the end of B, and then create a new vector with the initial point that of B and terminal point that of A':



The results are the same.

Exercise 1.6.2

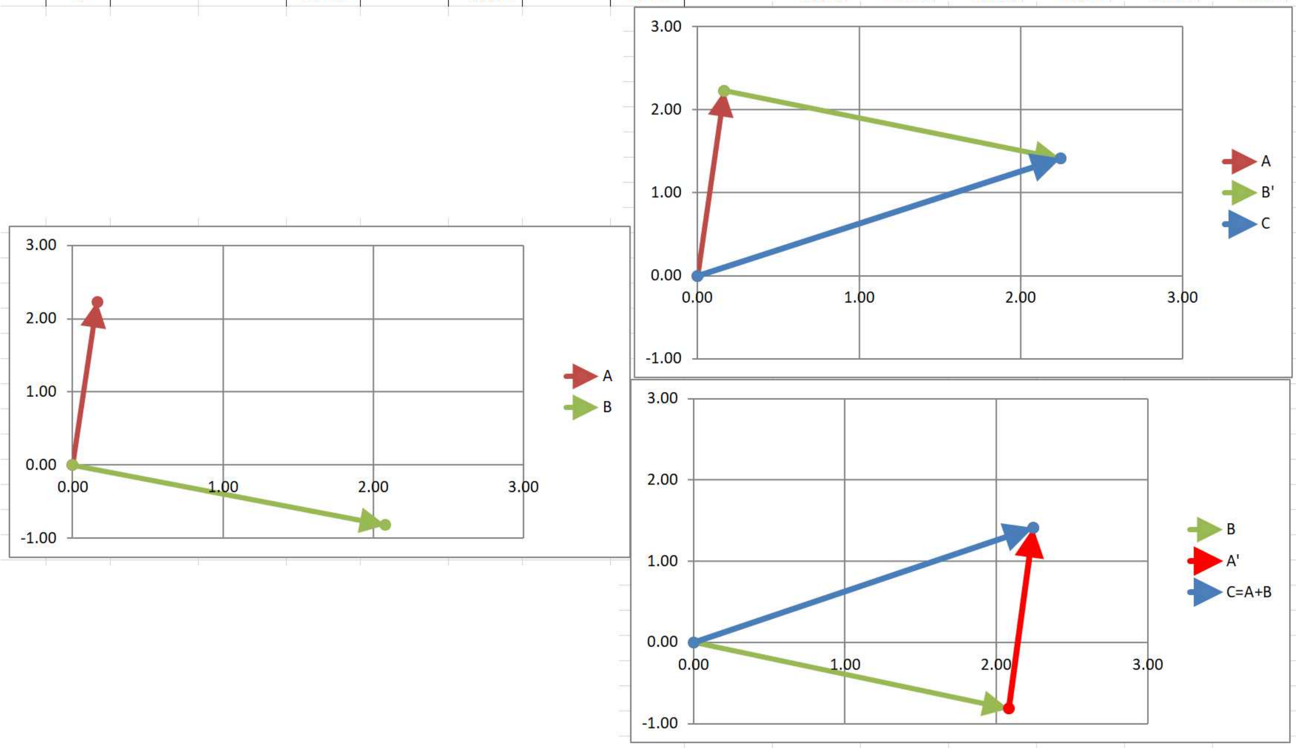
Explain why the results are the same. Hint: Similar triangles.

Therefore, we have the *Commutativity Property of Vector Addition*:

$$\boxed{A + B = B + A}$$

Now with the coordinate system present, the components of the vectors are combined as follows:

O	A	+	B	=	C	A		B		C	
0	0.16		2.08		2.25	0.00	0.16	0.00	2.08	0.00	2.25
0	2.23		-0.82		1.41	0.00	2.23	0.00	-0.82	0.00	1.41



Next, we know that we can ignore the parentheses when we are adding numbers:

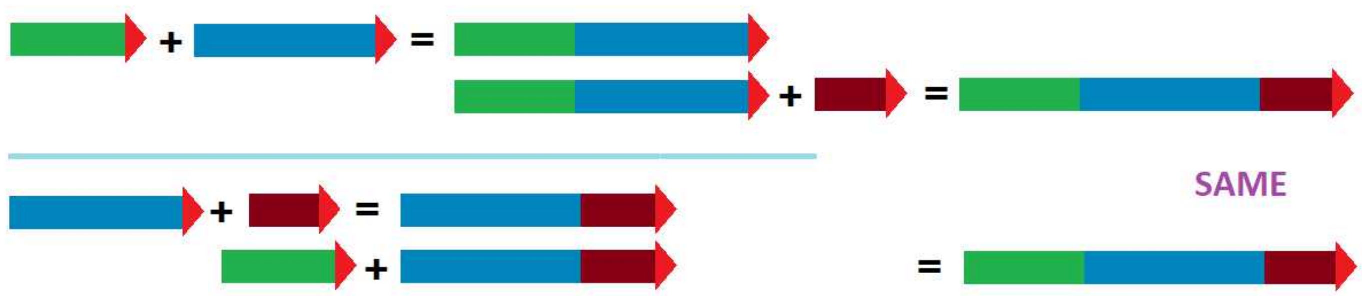
$$(1 + 2) + 3 = 1 + (2 + 3) = 1 + 2 + 3\,.$$

Identical is the *Associativity Property of Vector Addition*:

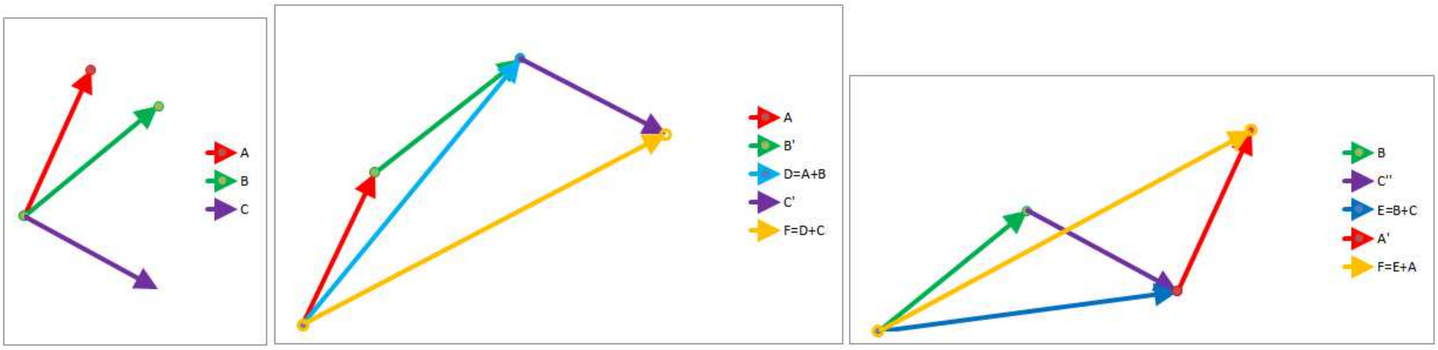
$$\boxed{A + (B + C) = (A + B) + C}$$

The order of addition doesn't matter!

This is the property of dimension 1:



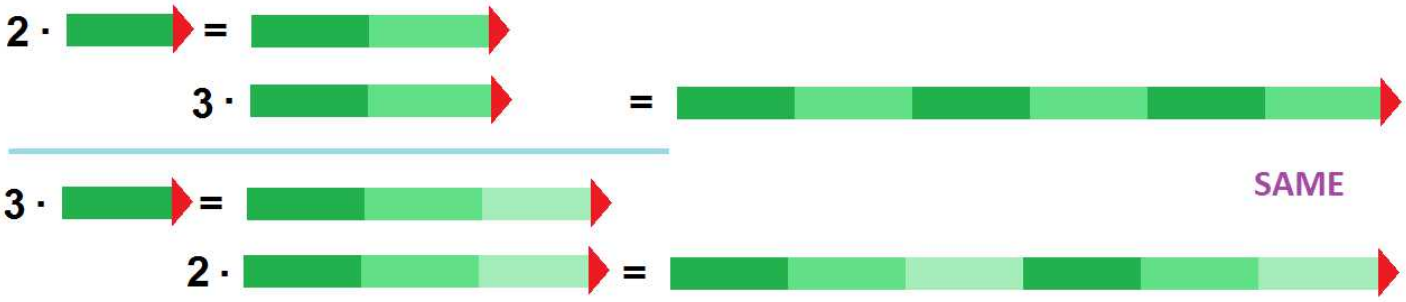
This is dimension 2:



Let's consider next how we can apply two (or more) scalar multiplications in a row. Given a vector A and real numbers a and b, we can create several new vectors:

- $B = aA$ from A and then $C = bB$ from B
- $D = bA$ from A and then $C = aD$ from D
- $C = (ab)A$ directly from A

The results are the same. For example, below we double, then triple:



The result is the same if we sextuple.

If we are to be precise, the missing symbol "$\cdot$" stands for two different things above:

- $B = a \cdot A$, scalar multiplication of a vector
- $D = b \cdot A$, scalar multiplication of a vector
- $C = (a \cdot b)A$, multiplication of numbers

Our conclusion is the *Associativity Property of Scalar Multiplication*:

$$\boxed{b(aA) = (ba)A}$$

Exercise 1.6.3

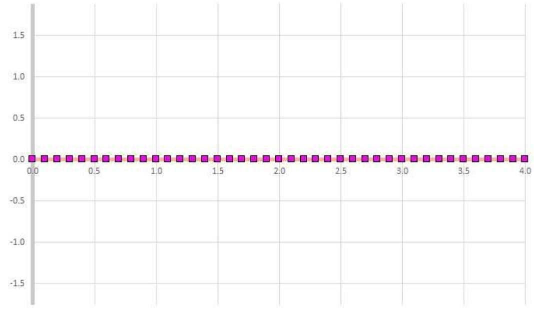
Provide an illustration for dimension 2.

There are some *special numbers*: 0 and 1. These formulas is what makes them special:

$$0 + x = x, \ 0 \cdot x = 0, \ 1 \cdot x = x\,.$$

Are there *special vectors*?

Consider vector addition. The *zero vector* is special. It has no magnitude nor direction and would have to be visualized as the dot O itself. This is what the zero vector is in $\mathbf{R}^{40}$:



There is 0, the real number, and then there is 0, the vector. The latter can mean no displacement, no motion (zero velocity), forces that cancel each other, etc. The two are related:

$$\boxed{0 \cdot A = 0}$$

This is a simple expression with a tricky algebraic meaning:

$$\text{number} \quad \cdot \quad \text{vector} \quad = \quad \text{vector}$$

Of course, the following holds for all vectors:

$$\boxed{A + 0 = A}$$

Now, scalar multiplication. Consider:

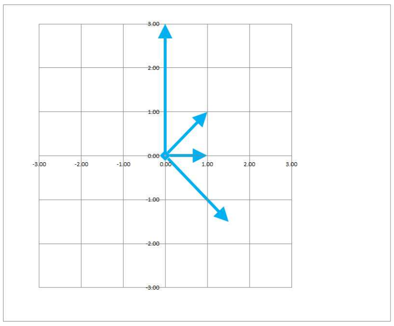
$$\boxed{1 \cdot A = A}$$

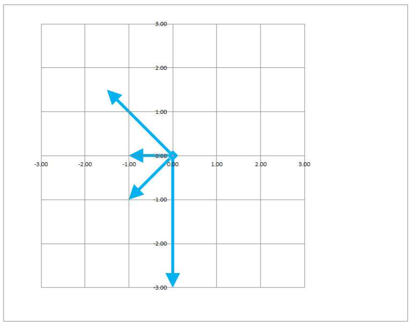
We have the same participants here as above:

$$\text{number} \quad \cdot \quad \text{vector} \quad = \quad \text{vector}$$

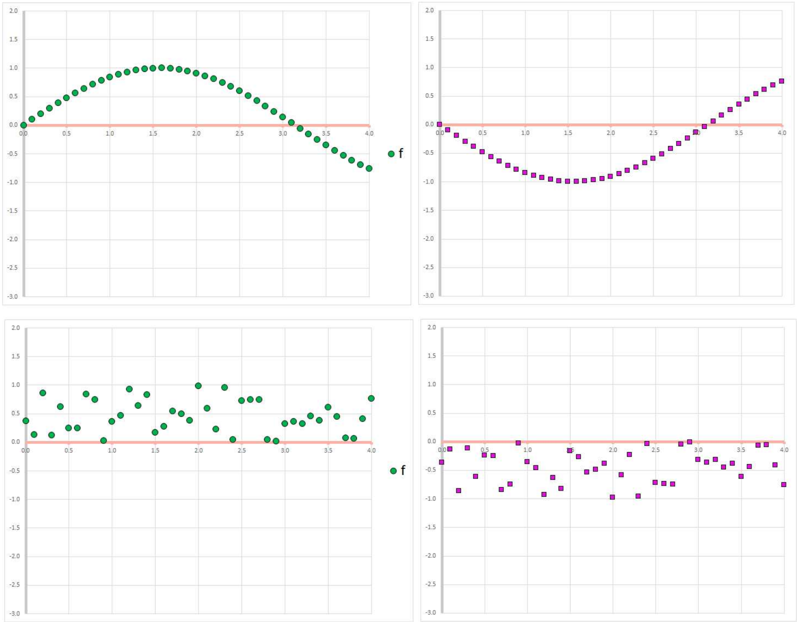
So, 1 remains special in vector algebra.

Next, since $PQ = -QP$, we have the *negative* $-A$ of a vector A, as the vector that goes in reverse of A. They are acquired by the central symmetry of the plane:





This is what happens in $\mathbf{R}^{40}$:



From the algebra, we also discover that

$$\boxed{-A = (-1)\cdot A}$$

As a summary, this is the complete list of rules one needs to carry out algebra with vectors:

Theorem 1.6.4: Axioms of Vector Space

The two operations – addition of two vectors and multiplication of a vector by a scalar – in $\mathbf{R}^n$ satisfy the following properties:

1. $X + Y = Y + X$ *for all* X *and* Y.
2. $X + (Y + Z) = (X + Y) + Z$ *for all* X, Y, *and* Z.
3. $X + 0 = X = 0 + X$ *for some vector* 0 *and all* X.
4. $X + (-X) = 0$ *for any* X *and some vector* $-X$.
5. $a(bX) = (ab)X$ *for all* X *and all scalars* a, b.
6. $1X = X$ *for all* X.
7. $a(X + Y) = aX + aY$ *for all* X *and* Y.
8. $(a + b)X = aX + bX$ *for all* X *and all scalars* a, b.

Taken together, these properties of vectors match the properties of numbers perfectly!

We put forward the following idea:

▶ *All manipulations of algebraic expressions that we have done with numbers are now allowed with vectors – as long as the expression itself makes sense.*

In other words, we just need to avoid operations that haven't been defined: no multiplication of vectors, no division of vectors, no adding numbers to vectors (of course!), etc.

All spaces of vectors, vector spaces if you like, we have seen so far have been only $\mathbf{R}^n$. Are there others?

Example 1.6.5: subsets and subspaces

Let's fix the last coordinate in $\mathbf{R}^n$ and look at the algebra:

$$\begin{array}{llllll} < & a_1 & a_2 & \dots & a_{n-1} & a_n > \\ + & & & & & \\ < & b_1 & b_2 & \dots & b_{n-1} & b_n > \\ \hline < & a_1+b_1 & a_2+b_2 & \dots & a_{n-1}+b_{n-1} & a_n+b_n > \end{array} \quad \to \quad \begin{array}{llllll} < & a_1 & a_2 & \dots & a_{n-1} & 0 > \\ + & & & & & \\ < & b_1 & b_2 & \dots & b_{n-1} & 0 > \\ \hline < & a_1+b_1 & a_2+b_2 & \dots & a_{n-1}+b_{n-1} & 0 > \end{array}$$

It works exactly the same!

How scalar multiplication works is also matched. Furthermore, we would anticipate that the eight properties in the theorem are satisfied.

Let's denote this set as follows:

$$\mathbf{R}^n_0 = \{< x_1, x_2, ..., x_{n-1}, 0 >\} \subset \mathbf{R}^n \,.$$

If we drop the redundant 0's, we realize that this is just a copy of $\mathbf{R}^{n-1}$:

$$\to \quad \begin{array}{lllll} < & a_1 & a_2 & ... & a_{n-1} > \\ + & & & & \\ < & b_1 & b_2 & ... & b_{n-1} > \\ \hline < & a_1+b_1 & a_2+b_2 & ... & a_{n-1}+b_{n-1} > \end{array}$$

Definition 1.6.6: vector space

Any set with two operations that satisfy the conclusions of the theorem is called a *a vector space*.

So, $\mathbf{R}^{n-1}_0$ is a vector space, a *subspace* of $\mathbf{R}^n$.

Every subset of $\mathbf{R}^n$ is subject to the algebraic operations of the ambient space. How do we determine when this is a vector space? We just need to make sure that the algebra makes sense:

Theorem 1.6.7: Subspaces

Suppose U is a subset of a vector space that satisfies:

1. *If X and Y belong to U, then so does $X+Y$.*
2. *If X belongs to U, then so does kX for any number k.*

Then U is a vector space.

Exercise 1.6.8

For the last example, show that setting the last coordinate to a non-zero number won't create a vector space.

Exercise 1.6.9

For the last example, show that setting the several coordinates to zero will create a vector space.

Exercise 1.6.10

Prove that a line through 0 on the plane is a vector space.

Exercise 1.6.11

Prove the theorem.

1.7. Convex, affine, and linear combinations of vectors

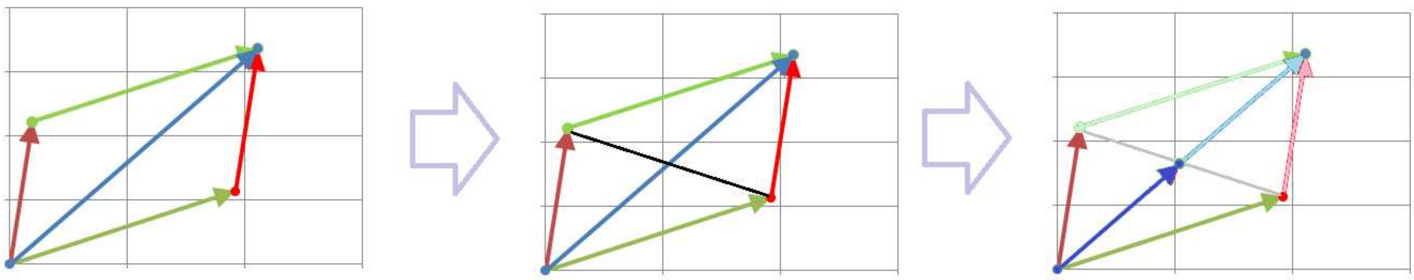
The average of two numbers is defined to be their half sum:

$$\frac{x+y}{2} = \frac{1}{2}x + \frac{1}{2}y \,.$$

Since the algebra of vectors mimics that of numbers, nothing stops us from defining the *average of vectors* U and V in the same manner:

$$\frac{1}{2}U + \frac{1}{2}V = \frac{1}{2}(U+V)\,.$$

It is a convenient concept illustrated below for dimension 2:



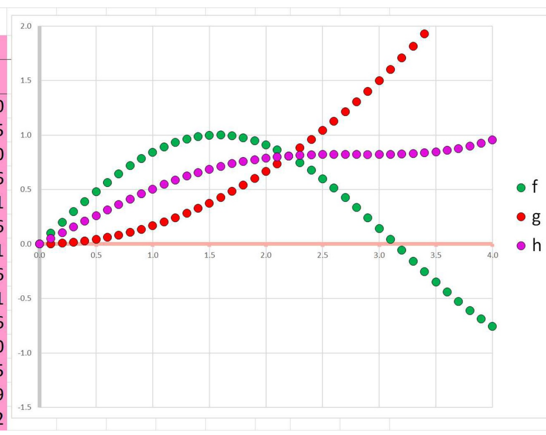
The average of two vectors is the vector that goes to the center of the parallelogram formed by the two.

Exercise 1.7.1

Prove the last statement.

In dimension 40, the graph of the average lies half-way (vertically) between the two:

n=	40				
h=	0.10		0.50	0.50	
		x	f	g	h
	1	0.0	0.00	0.00	0.00
	2	0.1	0.10	0.00	0.05
	3	0.2	0.20	0.01	0.10
	4	0.3	0.30	0.02	0.16
	5	0.4	0.39	0.03	0.21
	6	0.5	0.48	0.04	0.26
	7	0.6	0.56	0.06	0.31
	8	0.7	0.64	0.08	0.36
	9	0.8	0.72	0.11	0.41
	10	0.9	0.78	0.14	0.46
	11	1.0	0.84	0.17	0.50
	12	1.1	0.89	0.20	0.55
	13	1.2	0.93	0.24	0.59
	14	1.3	0.96	0.28	0.62



Let's take this idea one step further.

The weighted average of two numbers is defined to be a combination like this:

$$\alpha x + \beta y\,,$$

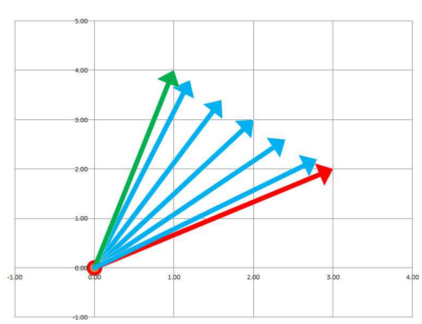
where $\alpha \geq 0$ and $\beta \geq 0$ add up to 1:

$$\alpha + \beta = 1\,.$$

Similarly, the *weighted average of vectors* U and V is defined to be the following:

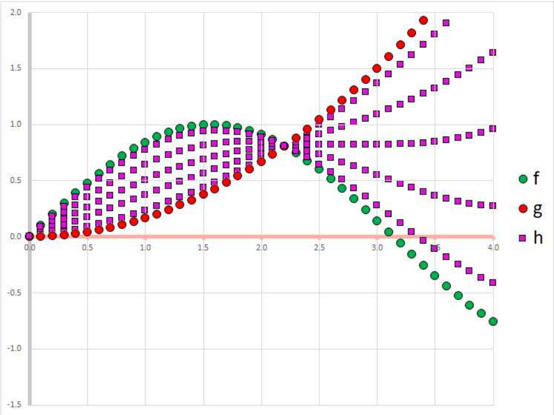
$$\alpha U + \beta V\,.$$

We can see in dimension 2 how one vector is gradually transformed into the other as *alpha* runs from 1 to 0 (and β from 0 to 1):



These intermediate stages are also called *convex combinations* of the two vectors. Together their ends form a line segment; it runs from the end of U to the end of V.

In dimension 40, we can also see a gradual transition from one vector to the other:

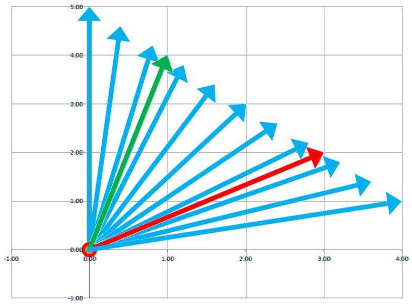


Now, we saw in dimension 2 that the average is also a straight segment between the two vectors. Of course, it is a straight segment between the two for dimension 3 or any dimension that we can visualize.

If we remove the restriction $\alpha \geq 0$ and $\beta \geq 0$, our combinations

$$\alpha U + \beta V$$

are called the *affine combinations* of U and V. Together, their ends form a whole line; it passes through the end of U and the end of V:



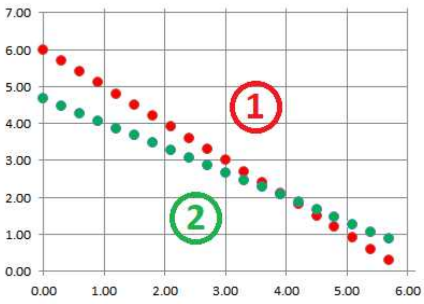
Let's recall the problem we have been using to illustrate many new ideas.

Problem: We are given the Kenyan coffee at \$2 per pound and the Colombian coffee at \$3 per pound. How much of each do you need to have 6 pounds of blend with the total price of \$14?

We let x be the weight of the Kenyan coffee and let y be the weight of the Colombian coffee. Then the total price of the blend is \$14. Therefore, we have a system:

$$\begin{array}{rcrl} x & + & y & = 6 \\ 2x & + & 3y & = 14 \end{array}$$

The solution to the system as presented initially had a clear *geometric* meaning. We thought of the two equations as equations about the coordinates of *points*, (x, y), in the plane. In fact, either equation is a representation of a line on the plane. Then the solution $(x, y) = (4, 2)$ is the point of their intersection:



The second interpretation was in terms of a *function* defined on the plane. A function $F : \mathbf{R}^2 \to \mathbf{R}^2$ is given by:

$$F(x, y) = (x + y, 2x + 3y)\,.$$

Then our solution is:

$$(x, y) = F^{-1}(6, 14)\,.$$

We now have a new interpretation – in terms of *vectors* in the plane.

We rewrite the system as a vector equation:

$$\begin{array}{ccccc} x & + & y & = 6 \\ 2x & + & 3y & = 14 \end{array} \quad \Longrightarrow \quad \begin{bmatrix} x & + & y \\ 2x & + & 3y \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

The first vector's components are computed via some algebra. We will try to interpret this algebra of numbers in terms of our algebra of vectors.

We split the vector up:

$$\begin{bmatrix} x & + & y \\ 2x & + & 3y \end{bmatrix} = \begin{bmatrix} x \\ 2x \end{bmatrix} + \begin{bmatrix} y \\ 3x \end{bmatrix}$$

We factor the repeated coefficients out:

$$\begin{bmatrix} x & + & y \\ 2x & + & 3y \end{bmatrix} = x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Our system has been reduced to a single *vector equation*:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

Let's analyze this equation and the problem it presents.

Given two vectors:

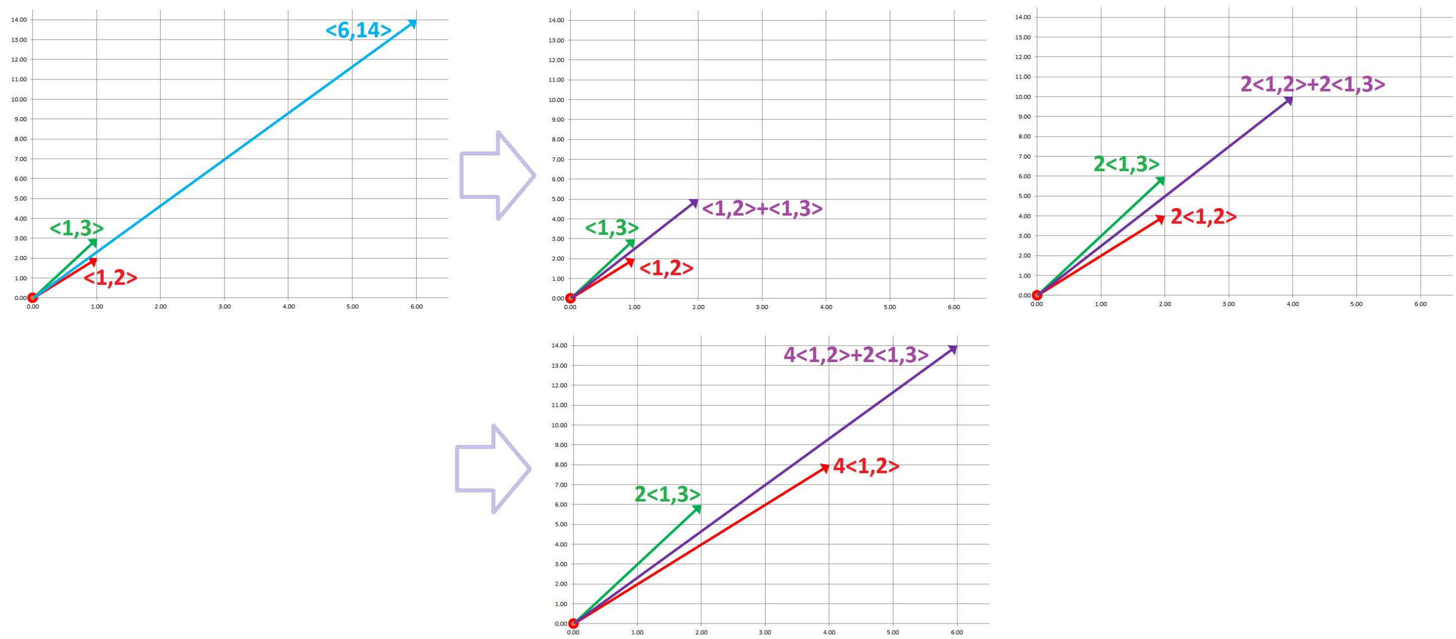
$$U = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

find two numbers x and y so that we have:

$$xU + yV = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

It may appear that we just need to represent the vector $< 6, 14 >$ as an *affine combination* of the vectors U and V. However, there is no restriction that x and y must add up to 1. We speak of *linear combinations*.

So, we need to find a way to *stretch* either of these two vectors so that their sum is the third vector. The setup is on the left followed by a trial-and-error:



Just adding the two vectors or adding their proportional multiples fails; it is clear that the angle can't match. Hypothetically, we go through all linear combinations of these two vectors to find one that is just right.

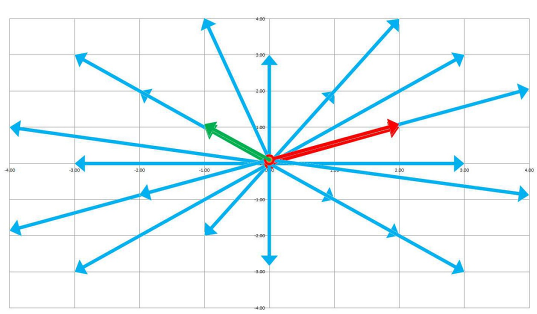
So, the new point of view on the problem of mixtures is different:

- ▶ Instead of the *locations*, we are after the *directions*.

In general, these are all linear combinations $\alpha U+\beta V$, in addition to the affine combinations:

$\alpha\backslash\beta$	...	-2	-1	0	1	2	...
...	...	...	...	...	...	...	...
-2	...	$-2U-2V$	$-2U-V$	$-2U$	$-2U+V$	$-2U+2V$	...
-1	...	$-U-2V$	$-U-V$	$-U$	$-U+V$	$-U+2V$	...
0	...	$-2V$	$-V$	0	V	$2V$	...
1	...	$U-2V$	$U-V$	U	$U+V$	$U+2V$	...
2	...	$2U-2V$	$2U-V$	$2U$	$2U+V$	$2U+2V$	...
...	...	...	...	...	...	...	...

These are the linear combinations with integer coefficients of $U=<2,1>$ and $V=<-1,1>$:



They seem to cover the whole plane.

Exercise 1.7.2

Is it true?

In summary, we have these three combinations:

Definition 1.7.3: linear, affine, and convex combinations

1. A *linear combination* of two vectors U and V is defined to be the following expression with any real numbers α and β called its coefficients:

$$\alpha U + \beta V\,.$$

2. An *affine combination* of two vectors U and V is defined to be their linear combination with coefficients α and β that add up to 1:

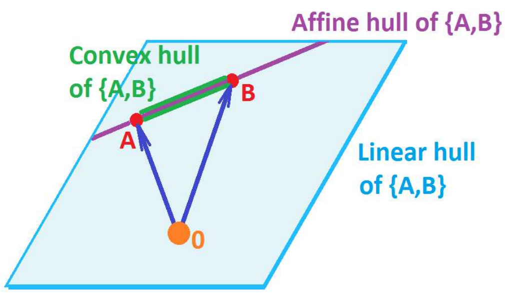
$$\alpha + \beta = 1\,.$$

3. A *convex combination* of two vectors U and V is defined to be their affine combination with coefficients α and β that are non-negative:

$$\alpha \geq 0,\ \beta \geq 0\,.$$

Example 1.7.4: hulls

These concepts have analogs for points. We define three *hulls*:



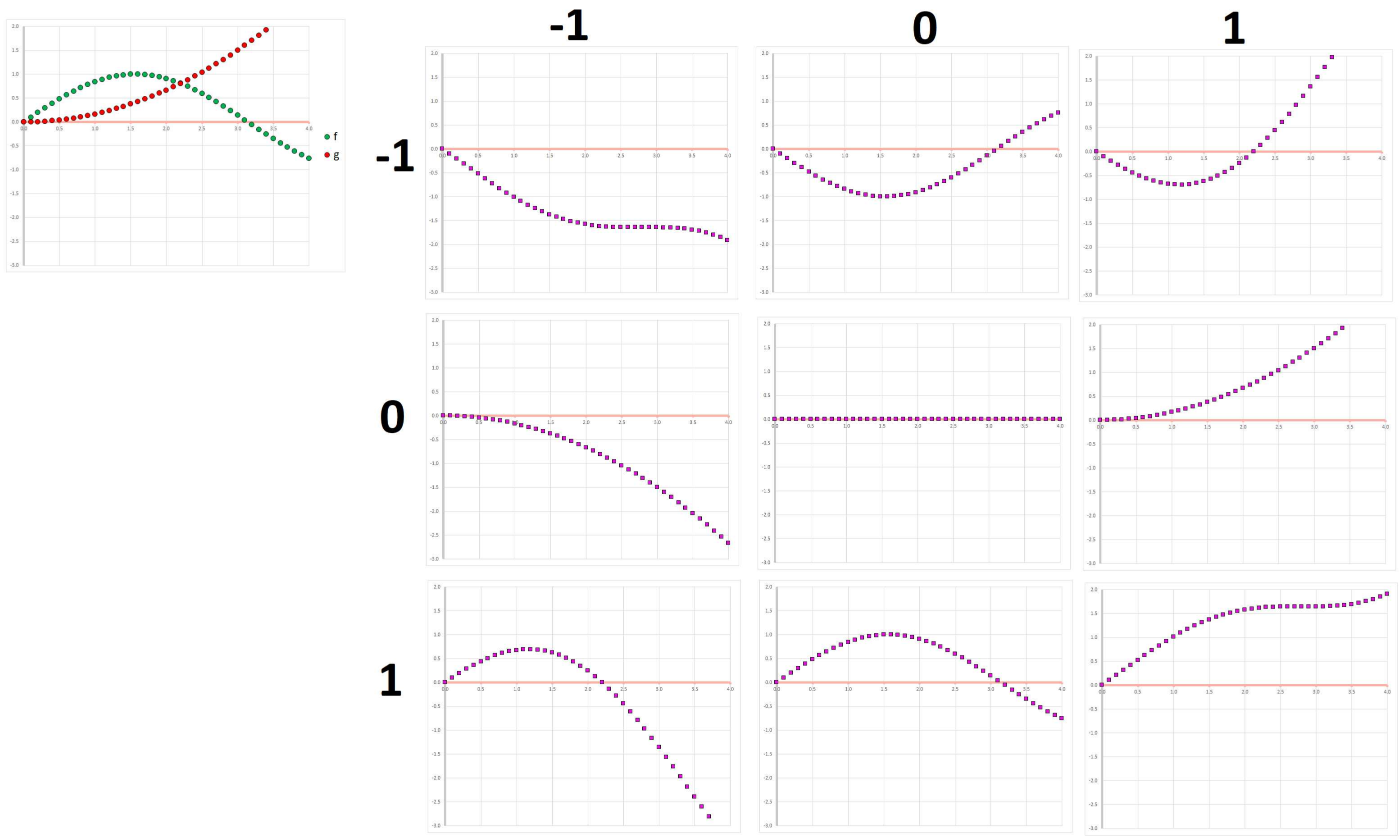
As you can see, the convex and the affine hulls are independent of the location of the origin.

Exercise 1.7.5

What about the linear hull?

In the general Euclidean space, we consider the results of all possible computations with the two algebraic operations that we have: vector addition and scalar multiplication.

This is what a few of them $(\alpha f + \beta g)$ look like in $\mathbf{R}^{40}$:



Exercise 1.7.6

Label these.

The next definition is an extension of this idea to an unlimited number of vectors:

Definition 1.7.7: linear combination

Suppose $V_1, ..., V_m$ are vectors in $\mathbf{R}^n$. Then, the *linear combination of the vectors* $V_1, ..., V_m$ *with coefficients* $r_1, ..., r_m$ is the following vector:

$$r_1 V_1 + ... + r_m V_m \,.$$

Then, the set of all linear combinations of a single vector is simply the set of its multiples (a line).

The set of all linear combinations of two vectors in the plane $\mathbf{R}^2$ is the whole plane, unless the two are multiples of each other.

Exercise 1.7.8

Prove the last statement.

Exercise 1.7.9

Finish the sentence: "The set of all linear combinations of three vectors in the 3-space is the whole 3-space, unless _________".

An important fact is the following:

Theorem 1.7.10: Linear Combination of Basis Vectors

Every vector in $\mathbf{R}^n$ *is a linear combination of the basis vectors:*

$$< a_1, ..., a_n >= a_1 e_1 + ... + a_n e_n \,.$$

Exercise 1.7.11

Prove the theorem.

Example 1.7.12: polynomials

A polynomial is a linear combination of the power functions:

$$a_0 + a_1x^1 + ... + a_nx^n .$$

In this sense, the space of all polynomials of degree up to n is indistinguishable from $\mathbf{R}^{n+1}$.

Exercise 1.7.13

Show that the multiples of a given vector in a vector space form a vector space.

Exercise 1.7.14

Show that the linear combinations of a given pair of vector in a vector space form a vector space.

Exercise 1.7.15

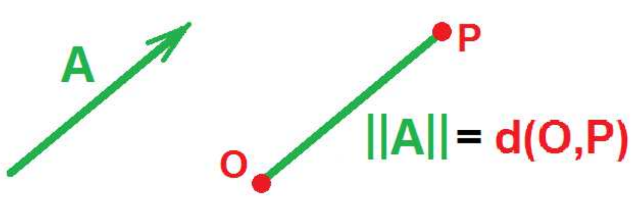
What is the next statement in this sequence?

1.8. *The magnitude of a vector*

A vector is a directed segment. Its attributes are, therefore, the direction and the magnitude. It may be hard to explain what *direction* means without referring, circularly, to vectors. That is why we look at the *magnitude* first:

▶ The magnitude of a vector is what's left of it when it's stripped off its direction.

When we interpret the vector as a displacement, we'd rather talk about its *length*. The meaning of this number is clear when the vector is given by two points, PQ. The length is the distance $d(P,Q)$ between them:



We intentionally make no reference to a Cartesian system:

Definition 1.8.1: magnitude of a vector

The *magnitude* or the *length of a vector* in $\mathbf{R}^n$ is defined to be the distance between its initial and terminal points:

$$\boxed{||PQ|| = d(P,Q)}$$

This number is also called the *norm* of the vector.

In particular, we have:

1. If $d(P,Q)$ stands for the Euclidean metric, $||PQ||$ is called the *Euclidean norm.*
2. If $d(P,Q)$ stands for the taxicab metric, $||PQ||$ is called the *taxicab norm.*

The notation resembles the *absolute value* and not by accident; they are the same in the 1-dimensional case, $\mathbf{R}$.

We also intentionally make no reference to a specific distance formula. The approach is as follows:

▶ We now look at each of the three *Axioms of Metric Space* and – using the above formula – translate it into a property of vectors.

First, the *Positivity*:

$$\boxed{d(P,Q) \geq 0; \quad \text{and} \quad d(P,Q) = 0 \ \text{ if and only if } P = Q}$$

We rewrite according to the definition above:

$$\boxed{||PQ|| \geq 0; \quad \text{and} \quad ||PQ|| = 0 \ \text{ if and only if } P = Q}$$

But the vector PP is just the zero vector! Therefore, we have this new form of the property:

$$\boxed{||A|| \geq 0; \quad \text{and} \quad ||A|| = 0 \ \text{ if and only if } \ A = 0}$$

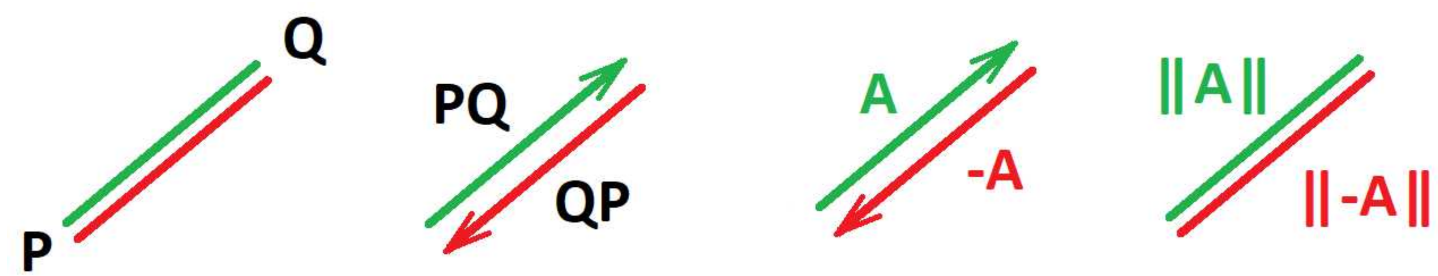
Second, the *Symmetry*:

$$\boxed{d(P,Q) = d(Q,P)}$$

We rewrite according to the definition above:

$$\boxed{||PQ|| = ||QP||}$$

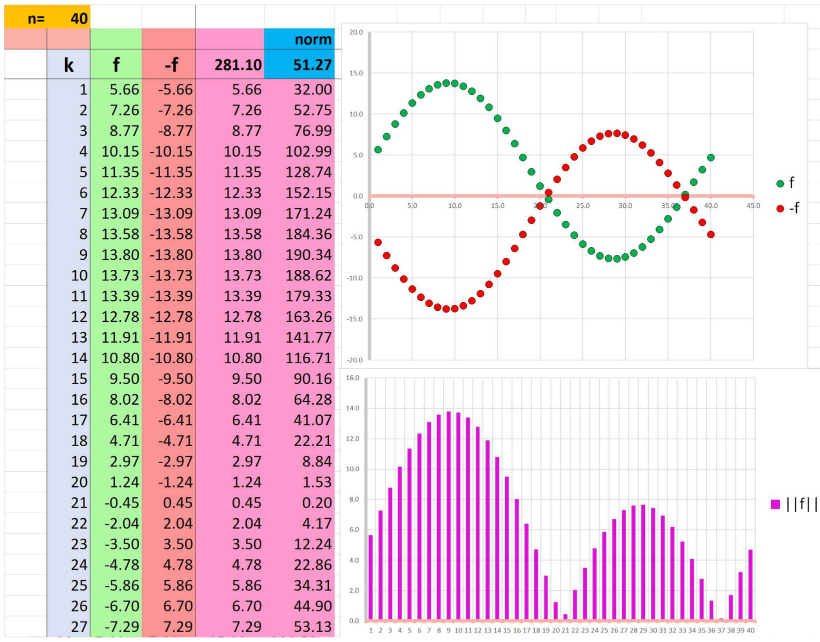
But the vector PQ is the negative of QP:



Therefore, we have this new form of the property:

$$\boxed{||A|| = ||-A||}$$

Its meaning is visualized for $\mathbf{R}^{40}$ below:



n=	40			
				norm
k	f	-f	281.10	51.27
1	5.66	-5.66	5.66	32.00
2	7.26	-7.26	7.26	52.75
3	8.77	-8.77	8.77	76.99
4	10.15	-10.15	10.15	102.99
5	11.35	-11.35	11.35	128.74
6	12.33	-12.33	12.33	152.15
7	13.09	-13.09	13.09	171.24
8	13.58	-13.58	13.58	184.36
9	13.80	-13.80	13.80	190.34
10	13.73	-13.73	13.73	188.62
11	13.39	-13.39	13.39	179.33
12	12.78	-12.78	12.78	163.26
13	11.91	-11.91	11.91	141.77
14	10.80	-10.80	10.80	116.71
15	9.50	-9.50	9.50	90.16
16	8.02	-8.02	8.02	64.28
17	6.41	-6.41	6.41	41.07
18	4.71	-4.71	4.71	22.21
19	2.97	-2.97	2.97	8.84
20	1.24	-1.24	1.24	1.53
21	-0.45	0.45	0.45	0.20
22	-2.04	2.04	2.04	4.17
23	-3.50	3.50	3.50	12.24
24	-4.78	4.78	4.78	22.86
25	-5.86	5.86	5.86	34.31
26	-6.70	6.70	6.70	44.90
27	-7.29	7.29	7.29	53.13

The norm is the purple area at the bottom.

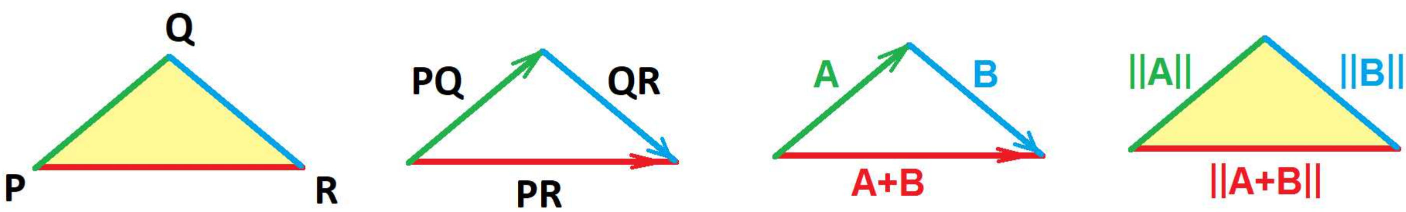
Third, the *Triangle Inequality*:

$$\boxed{d(P,Q) + d(Q,R) \geq d(P,R)}$$

We rewrite according to the definition above:

$$\boxed{||PQ|| + ||QR|| \geq ||PR||}$$

But $PQ + QR = PR$:



Therefore, we have this new form of the property:

$$\boxed{||A|| + ||B|| \geq ||A + B||}$$

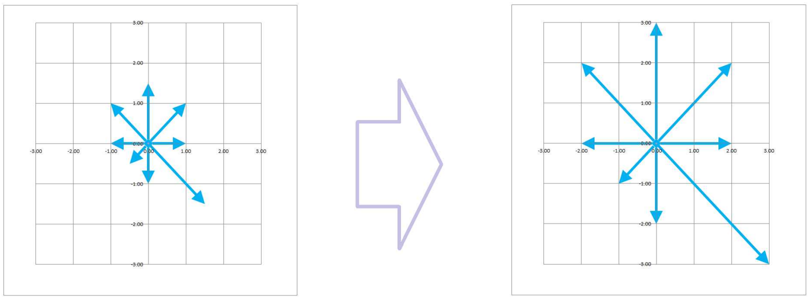
So, we have moved from the *geometry of points* to the *algebra of vectors*.

In summary:

1. The magnitude cannot be negative, and only the zero vector has a zero magnitude.
2. The magnitude of the negative of a vector is equal to that of the vector.
3. The magnitude of the sum of two vectors is larger than or equal to that of their sum.

That is how the magnitude interacts with *vector addition*. What about *scalar multiplication*? There is no corresponding property of distances.

All vectors double in length whether we multiply by 2 or -2, while their directions are preserved or flipped. When the direction doesn't matter, we just multiply the components by this stretching factor, $2 = |2| = |-2|$:



The result is another convenient property:

Theorem 1.8.2: Homogeneity of Norm

Both the Euclidean and the taxicab norms satisfy the following for any vector A and any scalar k:

$$\boxed{||k \cdot A|| = |k| \cdot ||A||}$$

Proof.

For the Euclidean norm in $\mathbf{R}^2$:

$$||k < a, b > || = || < ka, kb > || = \sqrt{(ka)^2 + (kb)^2} = |c| \cdot \sqrt{a^2 + b^2} = |k| \cdot || < a, b > || \,.$$

For the taxicab norm in $\mathbf{R}^2$:

$$||k < a, b > || = || < ka, kb > || = |ka| + |kb| = |k| \cdot |a| + |k| \cdot |b| = |k| \cdot (|a| + |b|) = k \cdot || < a, b > || \,.$$

Warning!

The *Symmetry* above is now redundant as it is incorporated into this new property:

$$-A = (-1)A \,.$$

These properties are applicable to all dimensions and are used to manipulate vector expressions.

We now turn around and ask:

- What's left of a vector when its magnitude is stripped off?

If we "remove" the magnitude from consideration, we are left with nothing but the direction. We can only say this:

- Vectors with the same direction are (positive) multiples of each other.

But what is the simplest vector among those?

To study directions, we limit our attention to some special vectors:

Definition 1.8.3: unit vector

Every vector with magnitude equal to 1 is called a *unit vector*:

$$||X|| = 1 \,.$$

We can make such a vector from any vector – "normalize" it – except 0, by dividing by its magnitude:

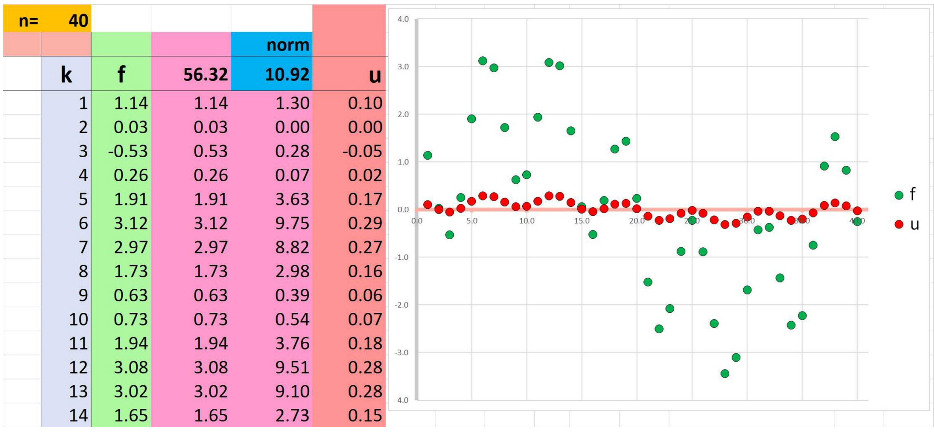
Theorem 1.8.4: Normalization of Vectors

For any vector $X \neq 0$*, the vector*

$$\boxed{Y = \frac{X}{||X||}}$$

is a unit vector.

It's simply a re-scaled version of the original. This is what it looks like in $\mathbf{R}^{40}$:

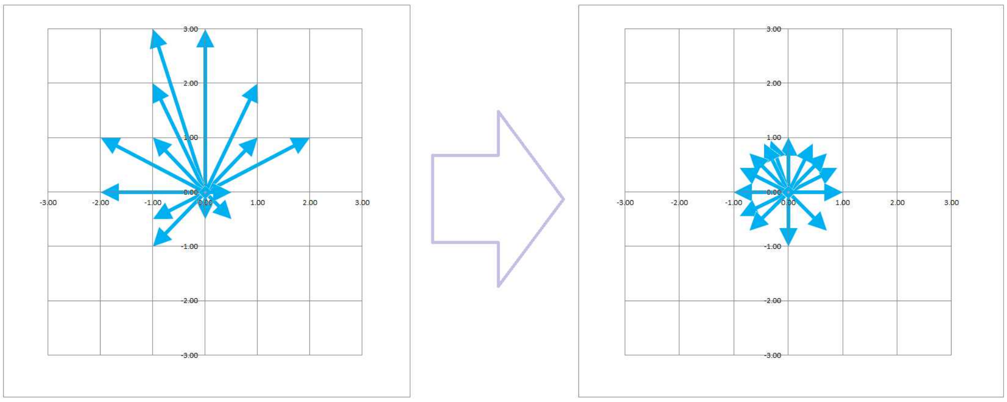


n=	40			
			norm	
k	f	56.32	10.92	u
1	1.14	1.14	1.30	0.10
2	0.03	0.03	0.00	0.00
3	-0.53	0.53	0.28	-0.05
4	0.26	0.26	0.07	0.02
5	1.91	1.91	3.63	0.17
6	3.12	3.12	9.75	0.29
7	2.97	2.97	8.82	0.27
8	1.73	1.73	2.98	0.16
9	0.63	0.63	0.39	0.06
10	0.73	0.73	0.54	0.07
11	1.94	1.94	3.76	0.18
12	3.08	3.08	9.51	0.28
13	3.02	3.02	9.10	0.28
14	1.65	1.65	2.73	0.15

Exercise 1.8.5

Prove the theorem.

The effect of normalization is that the vectors that are too long are shrunk and the ones that are too short are stretched – radially – toward the unit circle:



Unit vectors capture nothing but the direction:

Theorem 1.8.6: Multiples of Vectors

Suppose two vectors have equal or opposite unit vectors:

$$\frac{V}{||V||} = \pm \frac{W}{||W||} .$$

Then they are multiples of each other:

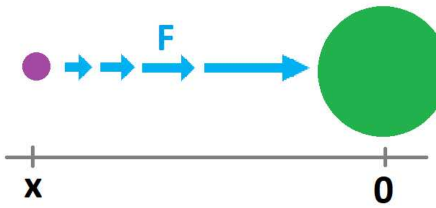
$$V = kW .$$

Exercise 1.8.7

Prove the theorem.

Example 1.8.8: Newton's Law of Gravity

Recall the law of gravity. Its force pulls two objects the harder the closer to each other they are:



The gravity is a function of two variables, x, y, that give the location, or better, it is a function of points, P, in $\mathbf{R}^2$ with real values:

$$f : \mathbf{R}^2 \to \mathbf{R}\,.$$

Algebraically, the law says that the force of gravity between two objects of masses M and m located at points O and P is given by:

$$\boxed{f(P) = G\frac{mM}{d(O,P)^2}}$$

Next, the law, in addition to the formula, includes the statement that the *force is directed from* P *to* O. This is implicitly the language of *vectors*: The direction of the force depends on the direction of the location vector. Let's sort this out.

Let's take care of the magnitudes first. We have two vectors:
- Gravity is a force and, therefore, a vector.
- The location of the second object is its displacement from the first and, therefore, a vector.

The function will have both vector inputs and vector outputs:

$$F : \mathbf{R}^2 \to \mathbf{R}^2\,.$$

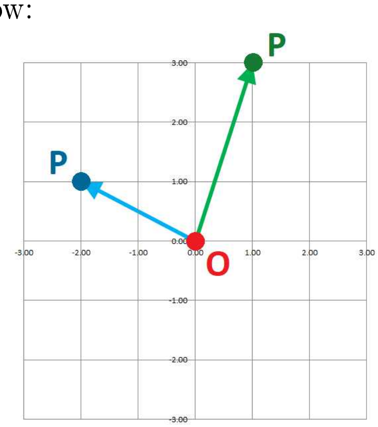
We place the origin O at the location of the first object (maybe the Sun). Then $OP = X$, and we can rewrite the formula as follows:

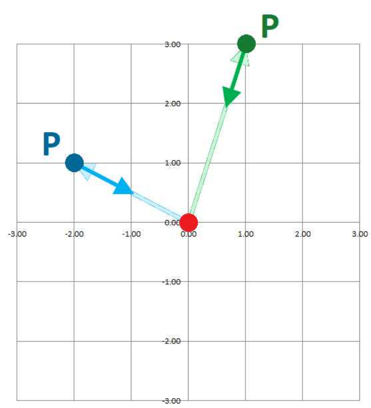
$$\boxed{||F(X)|| = G\frac{mM}{||X||^2}}$$

Next, what about the directions? Let's derive the *vector form* of the law. The law states:

▶ The force of gravity affecting either of the two objects is directed towards the other object.

We see this below:





In other words, F points in the opposite direction to X, i.e., it's direction is that of $-X$. Therefore, the unit vectors of $F(X)$ and $-X$ are equal:

$$\frac{F}{||F||} = -\frac{X}{||X||}\,.$$

Therefore, the vectors themselves are multiples of each other:

$$\boxed{F(X) = c(-X)}$$

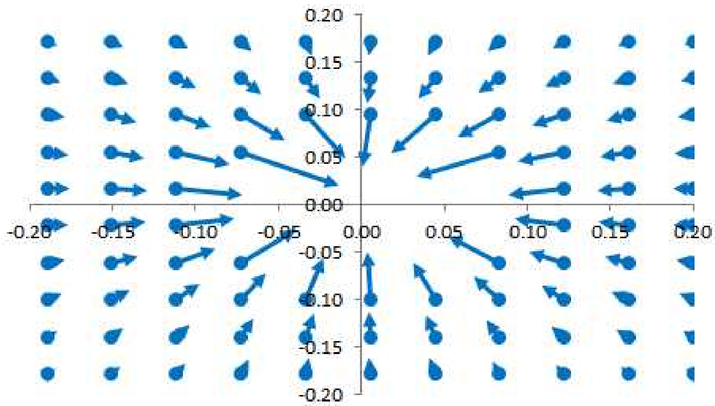
That's all we need except for the coefficient. We now use *Homogeneity* to find it:

$$G\frac{mM}{||X||^2} = ||F|| = |c| \cdot ||(-X)|| = |c| \cdot ||X||\,.$$

The final form is the following:

$$\boxed{F(X) = -G\frac{mM}{||X||^3}X}$$

Now that both the input and the output are 2-dimensional vectors, how do we visualize this kind of function? Even though this is just a (non-linear) transformation of $\mathbf{R}^2$ (or $\mathbf{R}^3$), there is a better way. First, we think of the input as a *point* and the output as a *vector* and then we attach the latter to the former. Below, we plot vector $F(X)$ starting at location X on the plane:



It is called a *vector field.*

Exercise 1.8.9

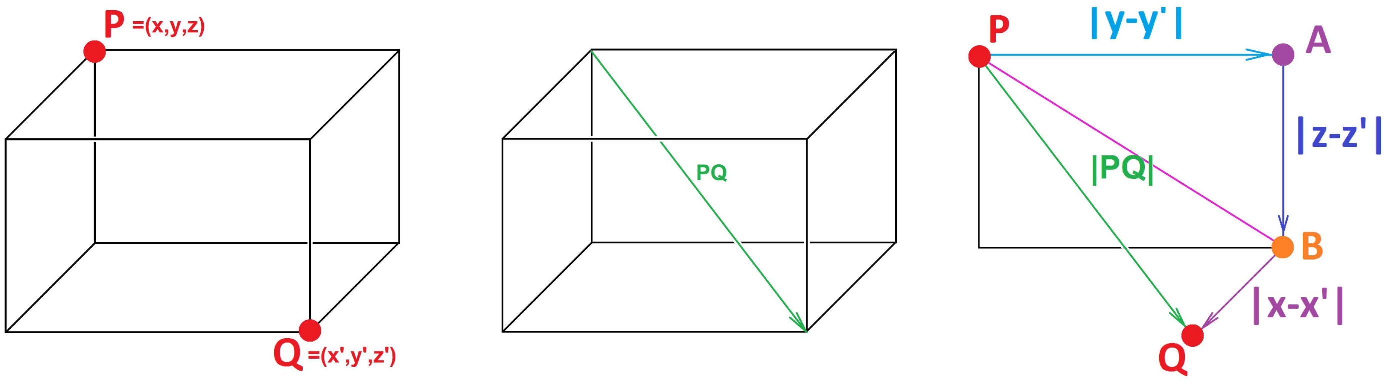
Suggest other examples of location-dependent forces.

These geometric properties have been justified following the familiar geometry of the "physical space" $\mathbf{R}^3$. However, they also serve as a *starting point* for a further development of linear algebra.

When a Cartesian system is provided, we have the *Euclidean metrics*, i.e., the distance between points P and Q in $\mathbf{R}^3$ with coordinates (x, y, z) and (x', y', z') respectively is

$$d(P, Q) = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}\,.$$

The three terms are recognized as the three components of the vector PQ:



If we apply this computation to a vector OP with $P = (a, b, c)$, we conclude:

$$|| < a, b, c > || = \sqrt{a^2 + b^2 + c^2}\,.$$

Meanwhile, the taxicab norm of this vector is

$$|| < a, b, c > || = |a| + |b| + |c|\,.$$

In general we have the following:

Theorem 1.8.10: Magnitude of Vector

Suppose we have a vector

$$A =< a_1, ..., a_n >$$

in $\mathbf{R}^n$. *Then we have:*

1. *The Euclidean norm of A is equal to*

$$\boxed{||A|| = \sqrt{a_1^2 + ... + a_n^2}}$$

2. *The taxicab norm of a vector A is equal to*

$$\boxed{||A|| = |a_1| + ... + |a_n|}$$

In the sigma notation, we have, respectively:

$$||A|| = \sqrt{\sum_{k=1}^{n} a_k^2}\,,$$

and

$$||A|| = \sum_{k=1}^{n} |a_k|\,.$$

As a summary, these are the properties of the magnitudes of vectors.

Theorem 1.8.11: Axioms of Normed Space

For any vectors A, B in $\mathbf{R}^n$ and any real k, the following properties are satisfied by both the Euclidean norm and the taxicab norm:

1. ***Positivity:*** $||A|| \geq 0$; *and* $||A|| = 0$ *if and only if* $A = 0$.
2. ***Triangle Inequality:*** $||A|| + ||B|| \geq ||A + B||$.
3. ***Homogeneity:*** $||k \cdot A|| = |k| \cdot ||A||$.

Exercise 1.8.12

Demonstrate that the formulas for the norms satisfy those three properties.

Example 1.8.13: investment portfolios

A portfolio of stocks can be subject to these operations. Assuming that there are only these 10 stocks available, all portfolios are vectors (or points) in $\mathbf{R}^{10}$:

vector

		$	
		A	squared
1	AGTK	20.0	400.0
2	AKAM	0.3	0.1
3	BCOR	5.0	25.0
4	BIDU	11.0	121.0
5	BRNW	12.0	144.0
6	CARB	15.0	225.0
7	CCIH	0.8	0.6
8	CCOI	0.0	0.0
9	JRJC	1.0	1.0
10	WIFI	23.0	529.0
	SUM	88.1	1445.7
	norm		38.0

Euclidean norm

Taxicab norm

The taxicab norm (yellow) is just the total value of the portfolio. The Euclidean norm is in pink.

Let's consider the "direction" of this portfolio. We normalize this vector by dividing by 88.1 for the taxicab norm and by 38.0 for the Euclidean norm:

Taxicab

		$	%
		A	weights
1	AGTK	20.0	0.23
2	AKAM	0.3	0.00
3	BCOR	5.0	0.06
4	BIDU	11.0	0.12
5	BRNW	12.0	0.14
6	CARB	15.0	0.17
7	CCIH	0.8	0.01
8	CCOI	0.0	0.00
9	JRJC	1.0	0.01
10	WIFI	23.0	0.26
	SUM	88.1	1.00

Euclidean

		$		
		A	squared	normalized
1	AGTK	20.0	400.0	0.53
2	AKAM	0.3	0.1	0.01
3	BCOR	5.0	25.0	0.13
4	BIDU	11.0	121.0	0.29
5	BRNW	12.0	144.0	0.32
6	CARB	15.0	225.0	0.39
7	CCIH	0.8	0.6	0.02
8	CCOI	0.0	0.0	0.00
9	JRJC	1.0	1.0	0.03
10	WIFI	23.0	529.0	0.60
	SUM		1445.7	
	norm		38.0	1.00

The former simply consists of the percentages of the stocks within the portfolio.

Warning!

It wouldn't make sense to have the norm of a portfolio of *non-homogeneous* items, such as commodities:

$$< 10000 \text{ tons of wheat}, 20000 \text{ barrels of oil}, ... > ,$$

or currencies:

$$< \$100000,\ ¥1000000,\ ... > .$$

Exercise 1.8.14

When is the norm equal to the sum of the components?

1.9. *Parametric curves*

Functions may process an input of any nature and produce an output of any nature.

We represent a function diagrammatically as a *black box* that processes the input and produces the output:

$$\begin{array}{ccccc} \text{input} & & \text{function} & & \text{output} \\ x & \mapsto & \boxed{\ f\ } & \mapsto & y \end{array}$$

CONVENTION

We will use the *upper case* letters for the functions the outputs of which are (or may be) multidimensional, such as points and vectors:

$$F,\ G,\ P,\ Q,\ \ldots$$

We will use the *lower case* letters for the functions with numerical outputs:

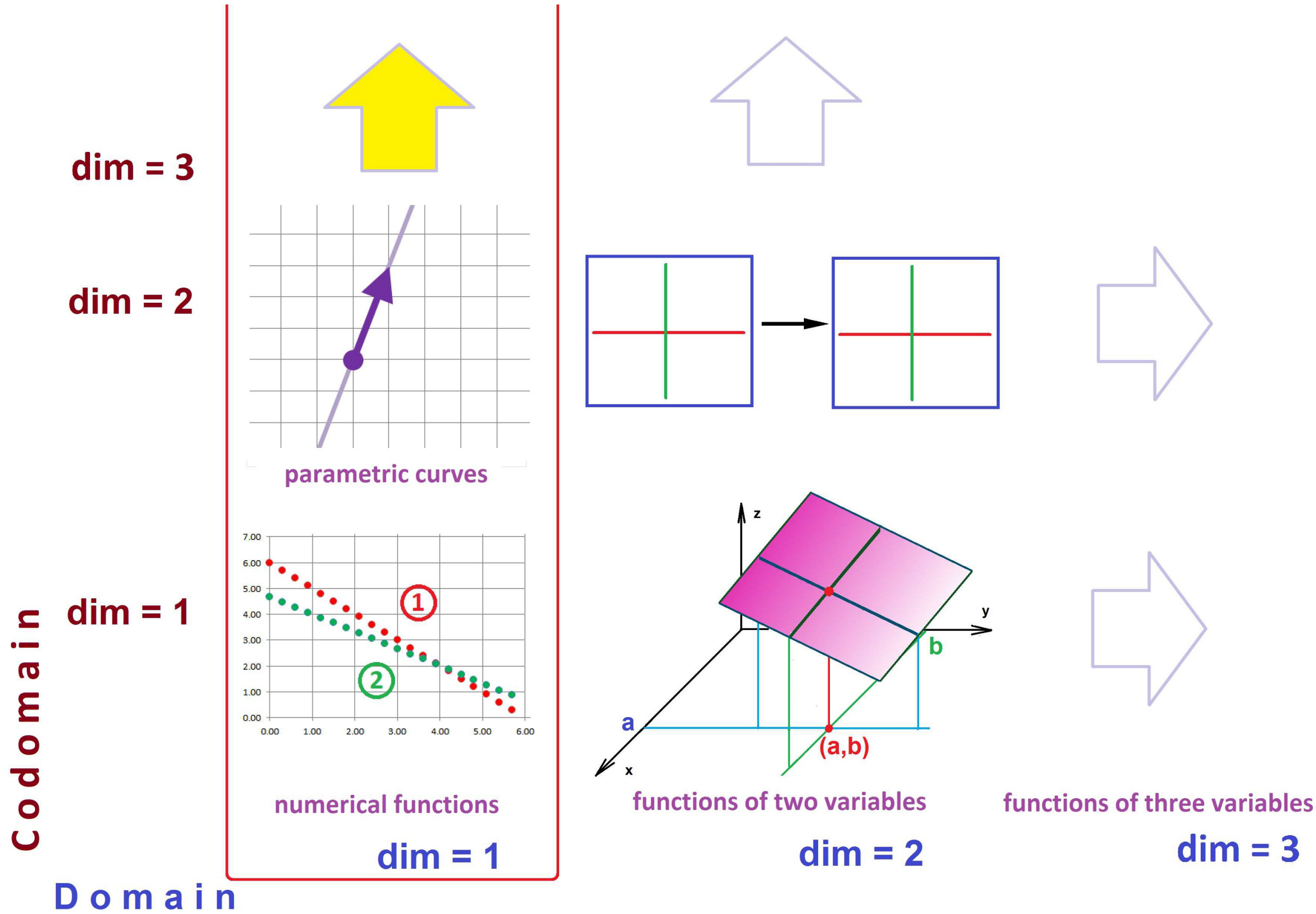
$$f,\ g,\ h,\ \ldots$$

Functions in multidimensional spaces take points or vectors as the input and produce points or vectors of various dimensions as the output. We can say that the input X is in $\mathbf{R}^n$ and the output $U = F(X)$ of X is in $\mathbf{R}^m$:

$$\begin{array}{cccc} F: & P & \mapsto & U \\ & \text{in } \mathbf{R}^n & & \text{in } \mathbf{R}^m \end{array}$$

Then, the domain of such a function is in $\mathbf{R}^n$ and the range (image) is in $\mathbf{R}^m$. The domain can be less than the whole space.

Below we illustrate the four (linear) possibilities for $n = 1, 2$ and $m = 1, 2$:

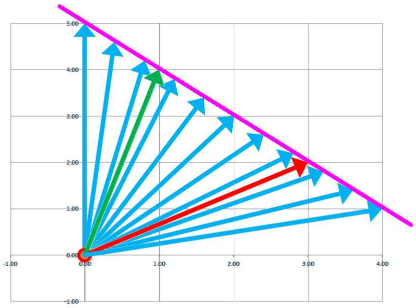


We will concentrate in this section on the first (infinite) column: parametric curves.

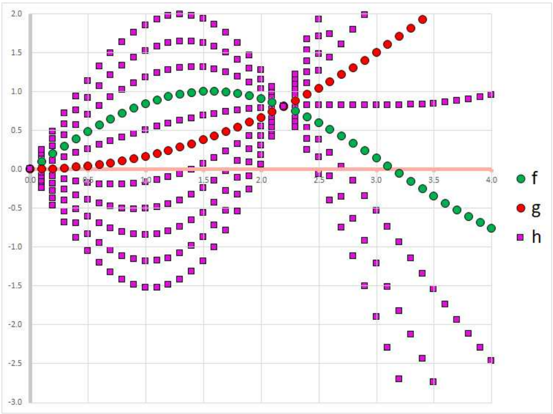
We will refer to as a *parametric curve* to

- any function of the real variable, i.e., the domain lies inside $\mathbf{R}$, and
- with its values in $\mathbf{R}^m$ for some $m = 1, 2, 3, \ldots$.

Recall from earlier in this chapter how straight lines appear as affine combinations of the two vectors on the plane:



And this is the line in $\mathbf{R}^{40}$ that passes through the two points shown in red and green:



In this section, we will limit ourselves to the interpretation of these functions via *motion*. The independent variable is the *time*, and the value is the *location*.

A *point* is the simplest curve. Such a curve with no motion is provided by a *constant function*.

A *straight line* is the second simplest curve.

We start with lines in $\mathbf{R}^2$. We already know how to represent straight lines on the plane:

1. The first method is the *slope-intercept form*:

$$y = mx + b\,.$$

   This method excludes the vertical lines! This is too limiting because in our study of curves, there are no preferred directions.

2. The second method is *implicit*:

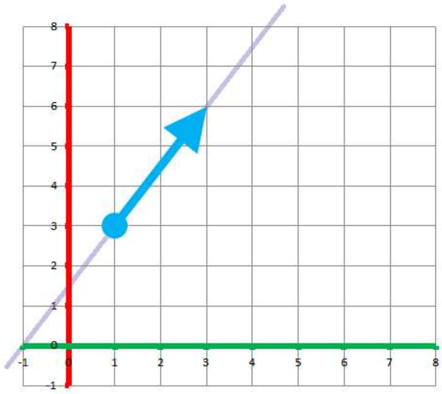
$$px + qy = r\,.$$

   The case of $p \neq 0,\ q = 0$ gives us a vertical line.

3. The third method is *parametric*. It has a dynamic interpretation (below).

Example 1.9.1: straight motion

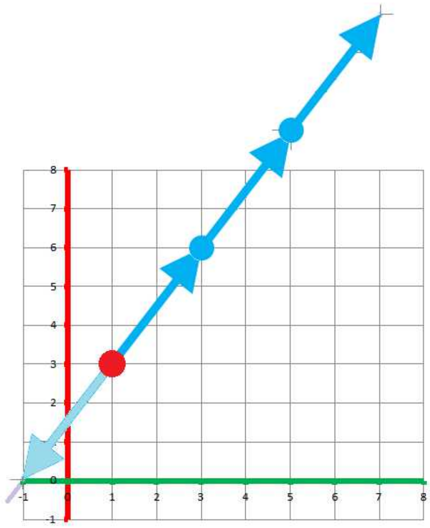
Suppose we would like to trace the line that starts at the point $(1,3)$ and proceeds in the direction of the vector $<2,3>$.



We use *motion* as a starting point and as well as a metaphor for parametric curves, as follows. We start moving:

- from the point $P_0 = (1, 3)$,
- under a constant velocity of $V =< 2, 3 >$.

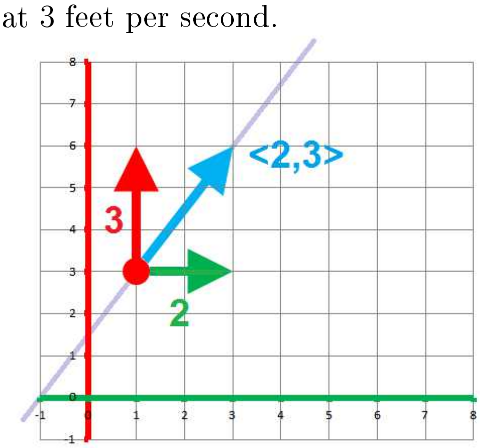
To get the rest of the path, we introduce another variable, time t. When $t = 1, 2, 3, ...$ is increasing incrementally, we have a sequence of locations on the plane:



For the negative t's, we go in the opposite direction.

Let's initially treat x and y separately:

- Horizontal: We move from 1 at 2 feet per second.
- Vertical: We move from 3 at 3 feet per second.



Let's find the formulas for the two. These are two consecutive locations:

$$x(0) = 1, \; x(1) = 3 \quad \text{and} \quad y(0) = 3, \; y(1) = 6\,.$$

The functions x and y must be linear:

$$x(t) = 1 + 2t \quad \text{and} \quad y(t) = 3 + 3t\,.$$

Combined, this is a parametric curve. Now, let's translate these formulas into the language of vectors.

In terms of vectors, if we are at point P now, we will be at point $P+V$ after one second. For example, we are at $P_1 = P_0 + V = (1,3) + < 2,3 >= (3,6)$ at time $t=1$. We define this function:

$$P : \mathbf{R} \to \mathbf{R}^2 \,.$$

It is made of two numerical functions:

$$P(t) = (x(t), y(t)) \,.$$

We already have two points on our parametric curve P:

$$P(0) = P_0 = (1,3) \quad \text{and} \quad P(1) = P_1 = (3,6) \,.$$

What is its formula?

We need to convert this to vectors:

$$x(t) = 1 + 2t,\ y(t) = 3 + 3t \,.$$

Let's assemble the two coordinate functions into one parametric curve:

$$P(t) = (x(t), y(t)) = (1 + 2t,\ 3 + 3t) \,.$$

This is still not good enough; we'd rather see the P_0 and V in the formula. We continue by using vector algebra:

$$\begin{array}{lll} P(t) & = (1+2t,\ 3+3t) & \text{We undo vector addition.} \\ & = (1,3) + < 2t,\ 3t > & \text{Then we undo scalar multiplication.} \\ & = (1,3) + t < 2,\ 3 > & \text{And finally we have the answer.} \\ & = P_0 + tV \,. & \end{array}$$

So, the four coefficients, of course, come from the specific numbers that give us P_0 and V.

Exercise 1.9.2

The line is not the graph of the function P but its _ _ _ _ _ .

We have discovered a vector representation of a straight uniform motion. The *location* P is given by:

$$P(t) = P_0 + tV \,,$$

where P_0 is the initial location and V is the (constant) *velocity*. Then tV is the *displacement*.

Warning!

One can, of course, move along a straight line at a *variable* velocity.

So, we have:

$$\boxed{\text{position at time } t \ = \ \text{initial position} \ + \ t \cdot \ \text{velocity}}$$

We used this approach for dimension 1; only the context has changed.

The pattern becomes clear. The line starting at the point (a, b) in the direction of the vector $< u, v >$ is represented parametrically as follows:

$$P(t) = (a, b) + t < u, v > \ .$$

Similarly for dimension 3, the line starting at the point (a, b, c) in the direction of the vector $< u, v, w >$ is represented as follows:

$$P(t) = (a, b, c) + t < u, v, w > \ .$$

And so on.

At the next level, we'd rather have no references to neither the dimension of the space nor the specific coordinates:

Definition 1.9.3: parametric curve of the uniform motion

Suppose P_0 is a point in $\mathbf{R}^m$ and V is a vector. Then the *parametric curve of the uniform motion through* P_0 *with the initial velocity of* V is the following:

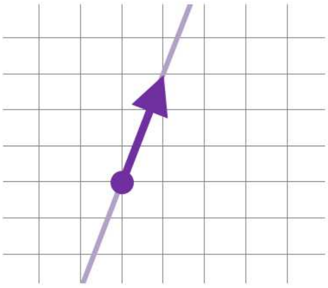
$$\boxed{P(t) = P_0 + tV}$$

Then, the *line through* P_0 *in the direction of* V is the path (image) of this parametric curve.

Stated for dimension $m = 1$, the definition produces the familiar point-slope form:

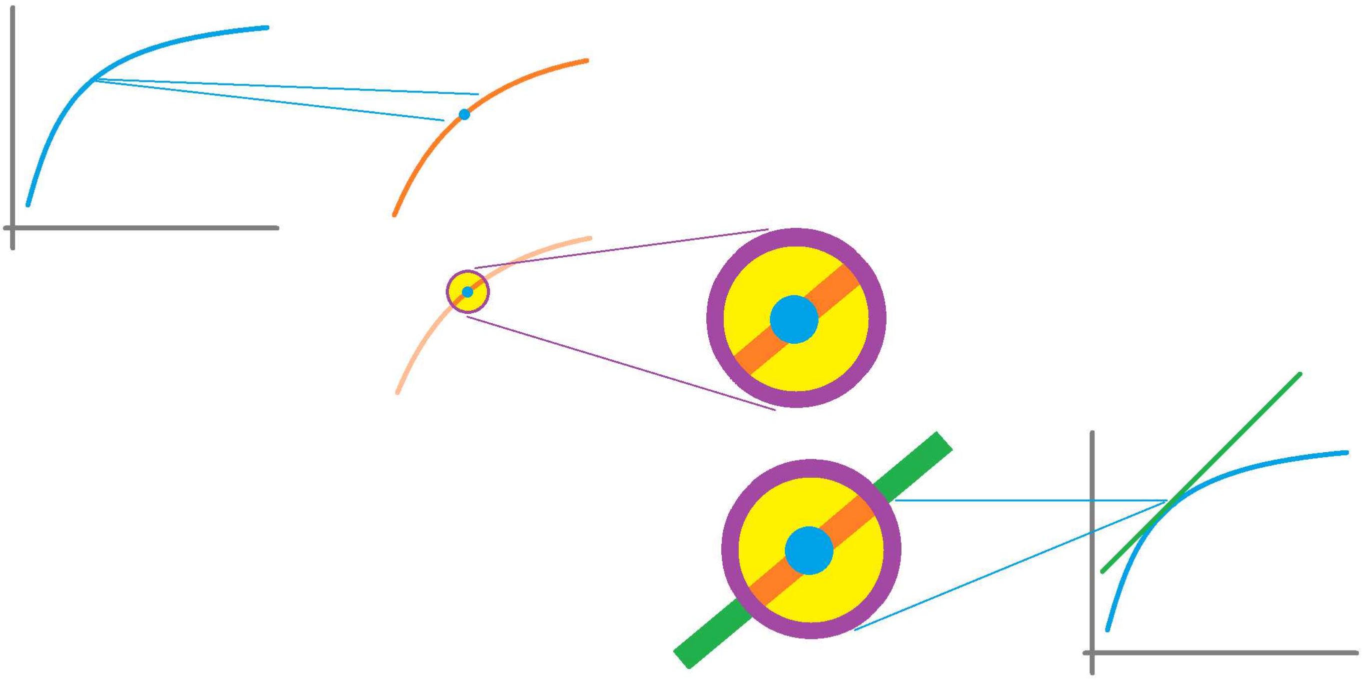
$$P(t) = P_0 + tV \, .$$

Indeed, P_0 is the y-intercept and V is the slope. The rate of change is a single number because the change is entirely within the y-axis. What has changed is the context as there are infinitely many directions in $\mathbf{R}^2$ for change:



That is why the change and the rate of change is a vector.

The importance of straight lines stems from the fact that, under common restrictions, *every* curve is likely to look like a straight line in the short term:

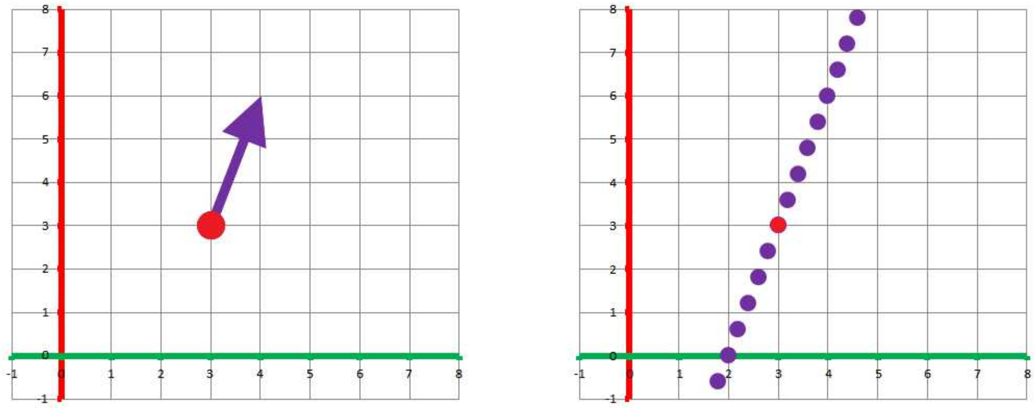


Example 1.9.4: recursive formulas

These are the recursive formulas that give the location as a function of time when the velocity is constant ($k = 0, 1, ...$):

$$\begin{array}{llll} x: & p_{k+1} & = p_k & +v\Delta t \\ y: & q_{k+1} & = q_k & +u\Delta t \end{array}$$

These points are plotted on the right for $p_0 = 3$, $q_0 = 3$, $v = 1$, $u = 3$, $\Delta t = 1/5$:



These quantities are now combined into points and vectors on the plane:

$$P_k = (p_k, q_k),\ V =< v, u > \ .$$

The equations take a vector form too:

$$P_{k+1} = P_k + V\Delta t\,.$$

Exercise 1.9.5

Consider the case when the velocity isn't constant.

Example 1.9.6: price dynamics

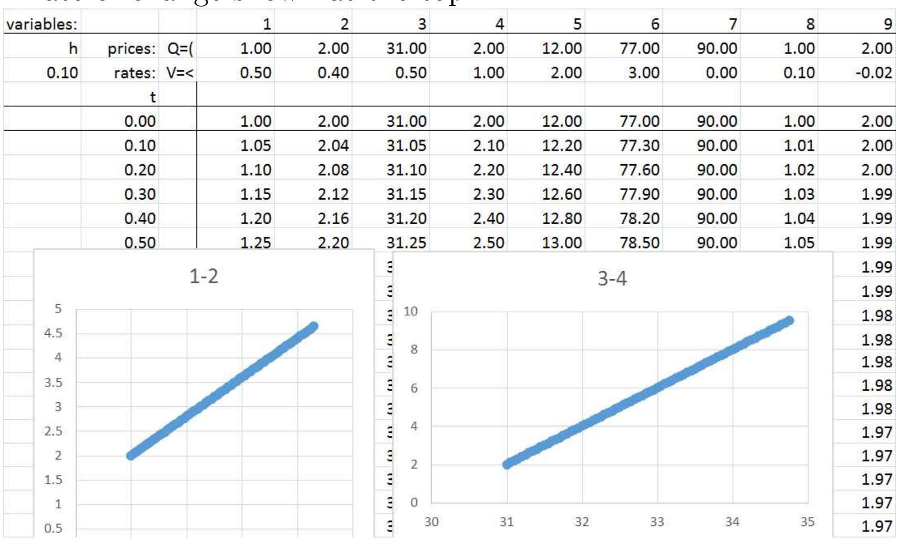
The definition applies to spaces of data. Suppose $\mathbf{R}^m$ is the space of prices (of stocks or commodities); we might have $m = 10,000$.

The prices recorded continuously or incrementally will produce a parametric curve and this curve might be a straight line. This happens when the prices are growing (or declining) *proportionally* but, possibly, at different rates.

Recursive formulas are especially easy to implement with a spreadsheet. In each column, we use the same formula for the kth price:

$$x_k(t+\Delta t) = x_k(t) + v_k\Delta t\,,$$

where v_k is the kth rate of change shown at the top:



variables:			1	2	3	4	5	6	7	8	9
h	prices:	Q=(	1.00	2.00	31.00	2.00	12.00	77.00	90.00	1.00	2.00
0.10	rates:	V=<	0.50	0.40	0.50	1.00	2.00	3.00	0.00	0.10	-0.02
	t										
	0.00		1.00	2.00	31.00	2.00	12.00	77.00	90.00	1.00	2.00
	0.10		1.05	2.04	31.05	2.10	12.20	77.30	90.00	1.01	2.00
	0.20		1.10	2.08	31.10	2.20	12.40	77.60	90.00	1.02	2.00
	0.30		1.15	2.12	31.15	2.30	12.60	77.90	90.00	1.03	1.99
	0.40		1.20	2.16	31.20	2.40	12.80	78.20	90.00	1.04	1.99
	0.50		1.25	2.20	31.25	2.50	13.00	78.50	90.00	1.05	1.99

The table gives us our curve. It lies in the $10,000$-dimensional space. Can we visualize such a curve in any way? Very imperfectly. We pick two columns at a time and plot that curve on the plane. Since these columns correspond to the axes, we are plotting a "shadow" (a projection) of our curve cast on the corresponding coordinate plane. They are all straight lines. A similar (short-term) dynamics may be exhibited by other data such as, for example, the vitals of a person:

1. body temperature
2. blood pressure
3. pulse (heart rate)
4. breathing rate

Exercise 1.9.7

Find a parametric representation of the line through two distinct points P and Q.

In the physical space, a straight line is followed by an object when there are no forces at play. Even a constant force leads to acceleration which may change the direction of the motion.

The advantage of the vector approach is that the choice of the coordinate system is no longer a concern!

Example 1.9.8: from relation to parametric

Suppose we have a line given by its relation:

$$y-3=2(x-1)\,.$$

What is its parametric representation?

Let's examine the equation. From its the slope-intercept form we derive:

1. 2 is the slope.
2. $(1,3)$ is the point.

So, let's just move

1. from the point $(1,3)$,
2. along the vector $<1,2>$ every second.

We have:

$$(x, y) = (1, 3) + t < 1, 2 > \ .$$

Exercise 1.9.9

What if we move faster?

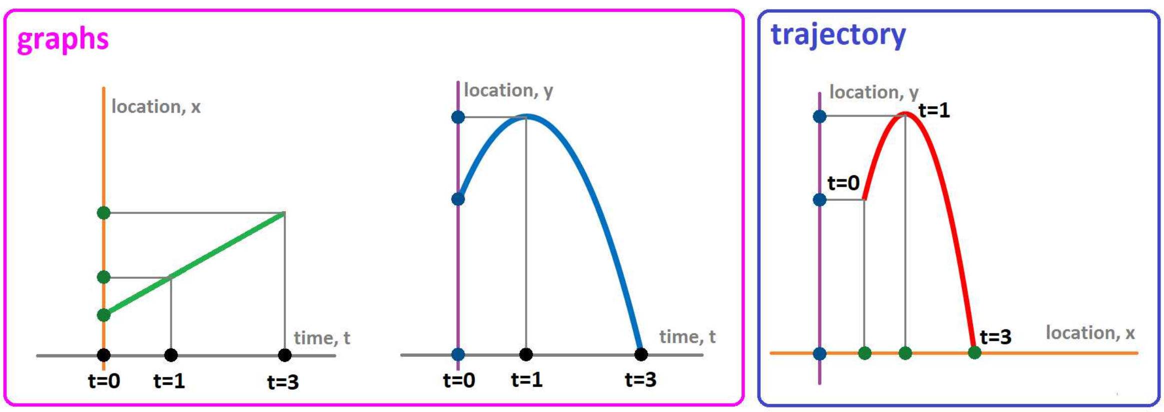
The example suggests a shortcut for $\mathbf{R}^2$:

$$\boxed{\text{slope } = \frac{\text{rise}}{\text{run}} \implies \text{direction } = < \text{run}, \text{rise} >}$$

Example 1.9.10: thrown ball

Let's review the dynamics of a thrown ball. A constant force causes the velocity to change linearly, just as the location in the last example. How does the location change this time?

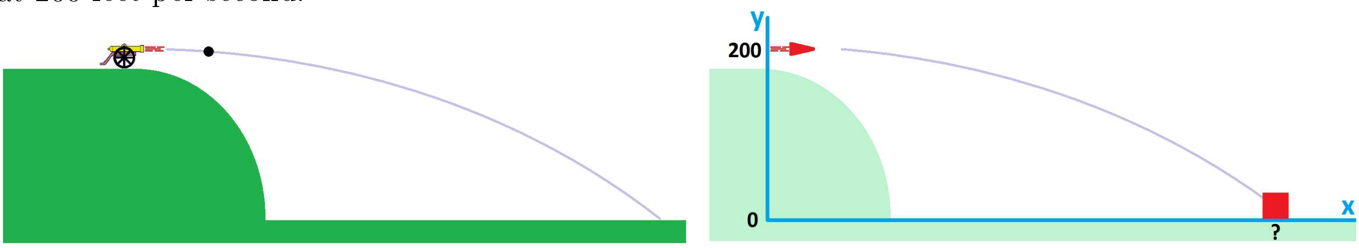
In the horizontal direction, as there is no force changing the velocity, the latter remains constant. Meanwhile, the vertical velocity is constantly changed by the gravity. The dependence of the height on the time is quadratic. The path of the ball will appear to an observer – from the right angle – as a curve:



A falling ball is subject to these accelerations, horizontal and vertical:

$$x: \ a_{k+1} = 0; \quad y: a_{k+1} = -g\,.$$

Now recall the setup considered previously: from a 200 feet elevation, a cannon is fired horizontally at 200 feet per second.



The initial conditions are:

- The initial location, $x: \ p_0 = 0$ and $y: \ p_0 = 200$.
- The initial velocity, $x: \ v_0 = 200$ and $y: \ v_0 = 0$.

Then we have two pairs of recursive equations – for the location in terms of the velocity and the velocity in terms of acceleration – independent of each other:

$$\begin{array}{llll} x: & v_{k+1} & = v_0, & \\ & & p_{k+1} & = p_k \quad +v_k\Delta t \\ y: & u_{k+1} & = v_k & -g\Delta t, \\ & q_{k+1} & = q_k & +u_k\Delta t \end{array}$$

These are the formulas in the vector notation:

$$\begin{array}{llll} V_{k+1} & = V_k & +A & \cdot\Delta t \\ P_{k+1} & = P_k & +V_{k+1} & \cdot\Delta t \end{array}$$

The advantage of the vector approach is that the choice of the coordinate system is no longer a concern!

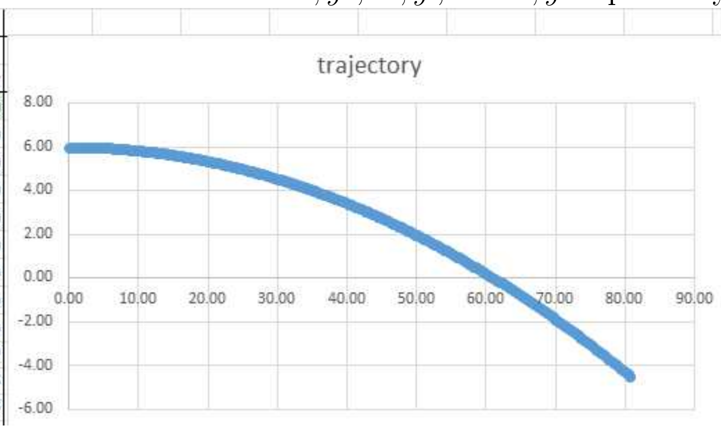
Example 1.9.11: recursive formulas

In dimension 2, for example, we don't have to align the x-axis with the direction of the throw and in dimension 3 we don't have to align the z-axis with the vertical direction. Nonetheless, let's start with former case. A 6-foot man throws – straight forward – a ball with the speed of 100 feet per second. If the throw is along the x-axis and the y-axis is vertical, we have:

$$A =< 0, -32 >,\ V_0 =< 0, 100 >,\ P_0 = (6, 0)\,.$$

This data goes into the first row of our table for the columns marked x'', y'', x', y', and x, y respectively.

0.003							
		A=X''		V=X'		X	
	t	x''	y''	x'	y'	x	y
	0.00	0.00	-32.00	100.00	0.00	0.00	6.00
	0.00	0.00	-32.00	100.00	-0.10	0.30	6.00
	0.01	0.00	-32.00	100.00	-0.19	0.60	6.00
	0.01	0.00	-32.00	100.00	-0.29	0.90	6.00
	0.01	0.00	-32.00	100.00	-0.38	1.20	6.00
	0.02	0.00	-32.00	100.00	-0.48	1.50	6.00
	0.02	0.00	-32.00	100.00	-0.58	1.80	5.99
	0.02	0.00	-32.00	100.00	-0.67	2.10	5.99
	0.02	0.00	-32.00	100.00	-0.77	2.40	5.99
	0.03	0.00	-32.00	100.00	-0.86	2.70	5.99
	0.03	0.00	-32.00	100.00	-0.96	3.00	5.98
	0.03	0.00	-32.00	100.00	-1.06	3.30	5.98



We apply the recursive formulas given above. In the spreadsheet,

- The velocity is computed from the velocity.
- The location is computed from the acceleration.

What is the difference of our *vectors* approach from our previous treatment of the flight of a ball? Instead of three columns for x'', x', x and then three columns for y'', y', y, one can see how the two components of acceleration, velocity, and location are combined into vectors contained in two columns each: x'', y'', then x', y', then x, y. The formula is almost the same as before:

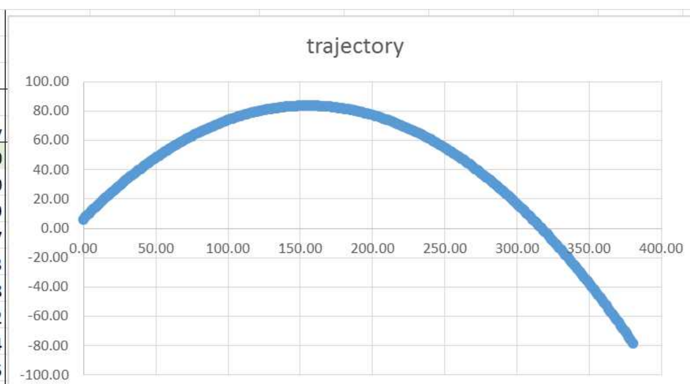
```
=R[-1]C+(RC[-2]-R[-1]C[-2])*R2C1
```

Next an angled throw... The only change is the vector of initial velocity:

$$A =< 0, -32 >,\ V_0 =< 100\cos\alpha,\ 100\sin\alpha >,\ P_0 = (6, 0)\,,$$

where α is the angle of the throw.

			speed=	100			
h=			angle=	45			
0.02							
		A=X''		V=X'		X	
	t	x''	y''	x'	y'	x	y
	0.00	0.00	-32.00	70.71	70.71	0.00	6.00
	0.02	0.00	-32.00	70.71	70.07	1.41	7.40
	0.04	0.00	-32.00	70.71	69.43	2.83	8.79
	0.06	0.00	-32.00	70.71	68.79	4.24	10.17
	0.08	0.00	-32.00	70.71	68.15	5.66	11.53
	0.10	0.00	-32.00	70.71	67.51	7.07	12.88
	0.12	0.00	-32.00	70.71	66.87	8.49	14.22
	0.14	0.00	-32.00	70.71	66.23	9.90	15.54
	0.16	0.00	-32.00	70.71	65.59	11.31	16.85
	0.18	0.00	-32.00	70.71	64.95	12.73	18.15
	0.20	0.00	-32.00	70.71	64.31	14.14	19.44



Example 1.9.12: continuous motion

Now the continuous case.

Starting with the physics,

$$\begin{cases} x'' &=& 0, \\ y'' &=& -g, \end{cases}$$

we integrate – coordinatewise – once:

$$\begin{cases} x' &=& & & v_x, & x'(0)=v_x & \text{is the initial horizontal velocity,} \\ y' &=& -gt &+& v_y, & y'(0)=v_y & \text{is the initial vertical velocity;} \end{cases}$$

and twice:

$$\begin{cases} x &=& & & v_xt &+& p_x, & x(0)=p_x & \text{is the initial horizontal position,} \\ y &=& -\frac{1}{2}gt^2 &+& v_yt &+& p_y, & y(0)=p_y & \text{is the initial vertical position.} \end{cases}$$

Thus, we have:

$$\begin{cases} \text{depth} &=& \text{initial depth} &+& \text{initial horizontal velocity} & \cdot\text{ time }, \\ \text{height} &=& \text{initial height} &+& \text{initial vertical velocity} & \cdot\text{ time} & -\frac{1}{2}g\cdot\text{ time }^2. \end{cases}$$

We take this solution to the next level by assembling these components into vectors just as in the last example.

$$\text{location} \quad = \quad \text{initial location} \quad + \quad \text{initial velocity} \quad \cdot\text{ time} \quad + \quad <0, -\frac{1}{2}g\cdot\text{ time }^2> .$$

The last term needs work. The zero represents the zero horizontal acceleration while $-g$ is the vertical acceleration. Then the last term is the acceleration times $\dfrac{t^2}{2}$. Algebraically, we have:

$$\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{c} p_x \\ p_y \end{array}\right] + \left[\begin{array}{c} v_x \\ v_y \end{array}\right] \cdot t + \left[\begin{array}{c} 0 \\ -g \end{array}\right] \cdot \frac{t^2}{2}.$$

The nature of the acceleration is irrelevant; we only need it to be constant.

Definition 1.9.13: parametric curve of uniformly accelerated motion

Suppose P_0 is a point in $\mathbf{R}^m$ and V_0, A are vectors. Then the *parametric curve of uniformly accelerated motion through P_0 with the initial velocity of V and acceleration A* is:

$$P(t) = P_0 + V_0\cdot t + A\cdot\frac{t^2}{2}\,.$$

We have an extra term, that disappears when $A = 0$, in comparison to the uniform motion. Just as in the 1-dimensional case, a constant acceleration produces a quadratic motion!

Exercise 1.9.14

Show that the path of this parametric curve is a parabola.

The values of a function represented by a parametric curve lie in $\mathbf{R}^m$ as *points* but can also be seen as *vectors*. For example, we can rewrite the familiar parametric curve of points:

$$P(t) = P_0 + V_0\cdot t + A\cdot\frac{t^2}{2}\,,$$

as one of vectors:

$$R(t) = R_0 + V_0\cdot t + A\cdot\frac{t^2}{2}\,.$$

Instead of passing through point P_0 it passes through the end point of vector $R_0 = OP_0$, which is the same thing. And, of course, the end of vector $R(t)$ is the point $P(t)$. The advantage of the latter approach is that it allows us to apply vector operations to the curves.

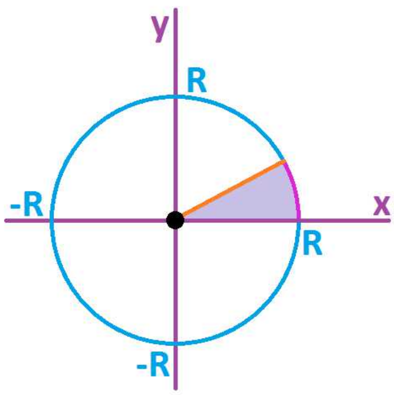
A more general approach to parametric curves, as well as their calculus, is presented in Chapter 2.

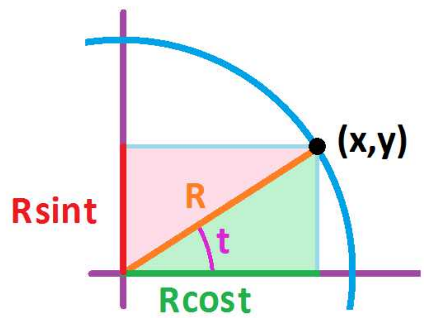
Example 1.9.15: circle transformed

Recall how we parametrized the unit circle using the angle as the parameter. Here, the x- and y-coordinates of a point at angle t is $\cos t$ and $\sin t$ respectively:

$$x = \cos t,\ y = \sin t\,.$$

The values of t may be the nodes of a partition of an interval such as $[0, 2\pi]$ or run through the whole interval.





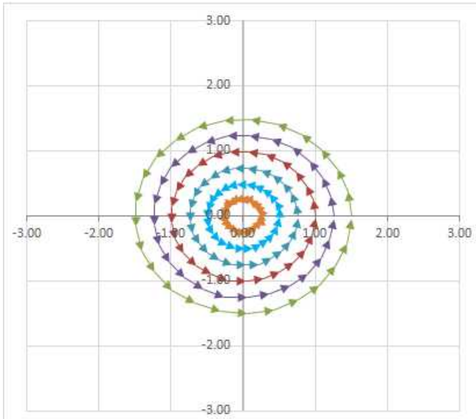
We can also look at this formula as a parametrization with respect to time. Then this is a record of motion with a constant speed or, in other words, a constant *angular* velocity. Now, this is the vector representation of this curve:

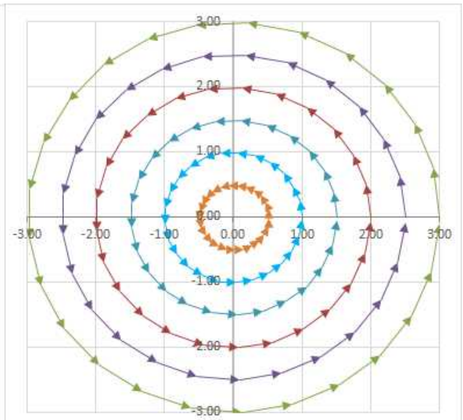
$$R(t) =< \cos t, \sin t > \ .$$

So, applying vector operations to this curve will give as new curves, just as in the 1-dimensional case (Chapter 1PC-4). For example, using scalar multiplication by 2 on all vectors means *stretching radially* the whole space. We then discover that the curve in the plane given by:

$$Q(t) = 2R(t) = 2 < \cos t, \sin t >\,,$$

is a parametric curve of the circle of radius 2.

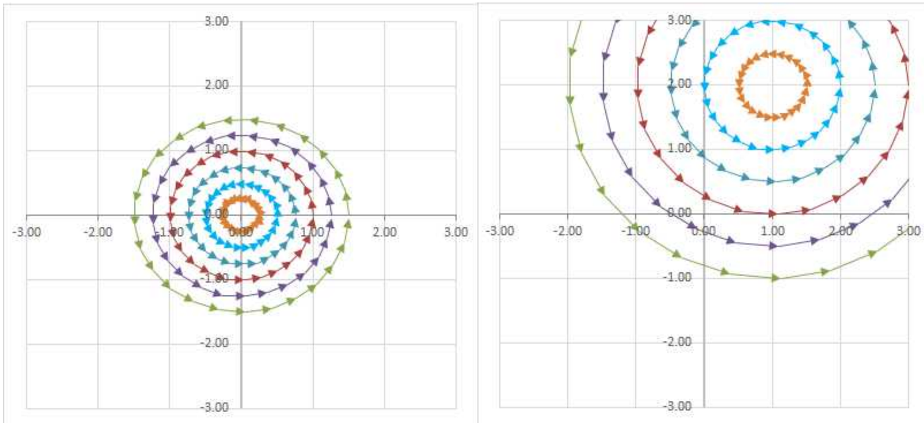




Similarly, using vector addition with $W =< 3, 1 >$ on all vectors means *shifting* the whole space by this vector. We then discover that the curve in the plane given by:

$$S(t) = W + 2R(t) = (1, 2) + 2 < \cos t, \sin t >\,,$$

is a parametric curve of the circle of radius 2 centered at $(1, 2)$.



And so on with other *transformations* of the plane.

1.10. The angles between vectors; the dot product

Recall that a Cartesian system pre-measures the space $\mathbf{R}^n$ so that we can do *analytic geometry*:

> ▶ Using the coordinates of points and the components of vectors, we compute distances and angles.

In this chapter, we applied this idea to the distances between points and, therefore, to the magnitudes of vectors. What about the *angles*? Let's first review what we did for dimensions 1 and 2.

Dimension 1 first.

What is the difference between the vectors OP and OQ (P, Q are not equal to O) represented in terms of their components x and x'? There can be only two possibilities:

- If P and Q are on the same side of O then the *directions are the same*,
- If P and Q are on the opposite sides of O then the *directions are the opposite*.

Then the theorem about the *directions for dimension* 1 is stated as follows:The angle between the vectors OP and OQ with components $x \neq 0$ and $x' \neq 0$ is

- 0 when $x \cdot x' > 0$; and
- π when $x \cdot x' < 0$.

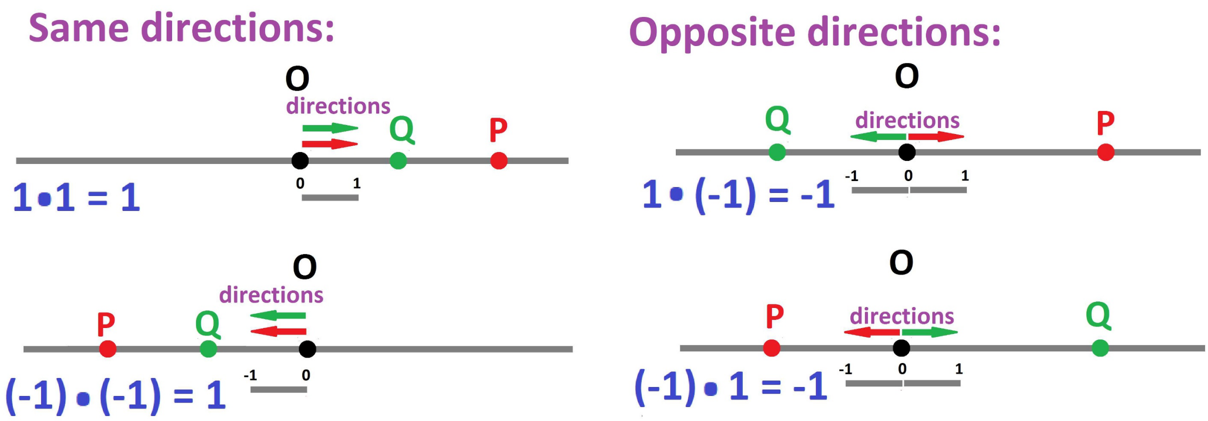
However, we have made some progress since we faced this task. Mainly, it is this realization:

> ▶ The direction of the vector *is* its normalization, a unit vector.

Indeed, only the directions of the vectors matter and not the sizes! We can then make the same statement but about the unit vectors:

$$\frac{x}{|x|} \quad \text{and} \quad \frac{x'}{|x'|}\,.$$

The advantage is that they can only take two possible values, 1 and -1, the positive direction and the negative direction. And so does their product:



We can then restate the result: The angle between vectors OP and OQ with components $x \neq 0$ and $x' \neq 0$ is the following:

normalization of x	$\cdot$	normalization of y		product	angle
$\dfrac{x}{\lvert x\rvert}$	$\cdot$	$\dfrac{x'}{\lvert x'\rvert}$	$=$	1	0
$\dfrac{x}{\lvert x\rvert}$	$\cdot$	$\dfrac{x'}{\lvert x'\rvert}$	$=$	-1	π

Matching these four numbers,

$$0 \mapsto 1 \quad \text{and} \quad \pi \mapsto -1\,,$$

we realize that this is the cosine:

$$\cos 0 = 1 \quad \text{and} \quad \cos \pi = -1\,.$$

We then have a new version of our theorem:

Theorem 1.10.1: Angles for Dimension 1

If θ is the angle between vectors OP and OQ in $\mathbf{R}$ with components $x \neq 0$ and $x' \neq 0$ then

$$\boxed{\cos\theta = \frac{x}{|x|} \cdot \frac{x'}{|x'|}}$$

Now the coordinate system for dimension 2.

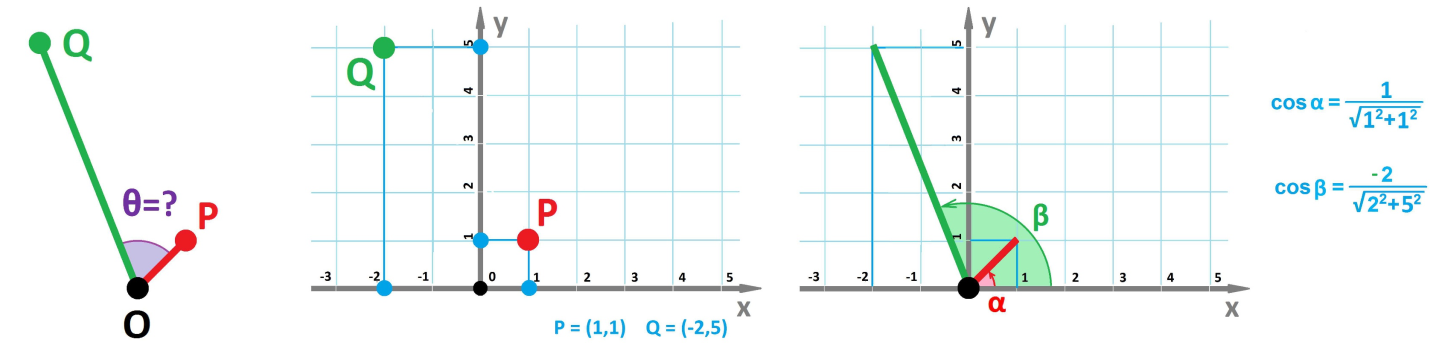
What is the difference between the directions from the origin O toward points P and Q (other than O) represented in terms of their coordinates (x, y) and (x', y')? We are talking about *angle* between the two non-zero vectors:

$$OP =< x, y > \quad \text{and} \quad OQ =< x', y' > \ .$$

Warning!

If one of the vectors is zero, there is no angle because there is no direction.

We know how to find the angles with the x-axis, α and β:



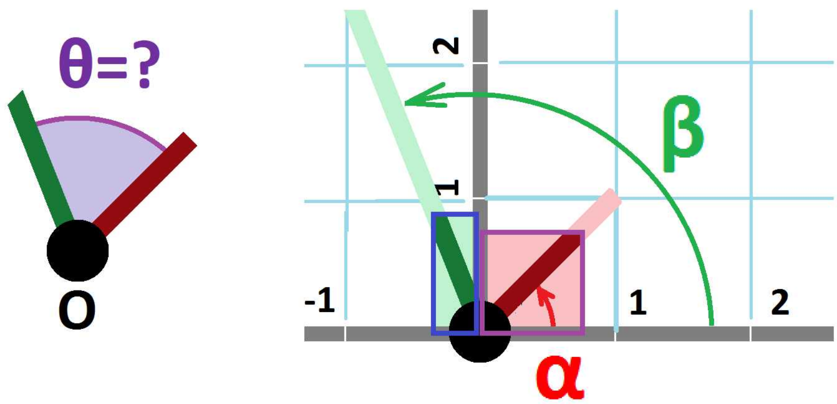
The angle we are looking for is:

$$\theta = \widehat{QOP} = \beta - \alpha\,.$$

The cosine of this angle can be found from the trigonometric functions of these two angles according to the following formula:

$$\cos\theta = \cos(\beta - \alpha) = \cos\beta\cos\alpha + \sin\beta\sin\alpha\,.$$

Let's exclude the magnitudes of the vectors from consideration:



Instead of the original vectors OP and OQ, we look their *normalizations*, U and V, respectively:

$$\begin{array}{lll} OP =< x,y > & \Longrightarrow\ U = \dfrac{< x,y >}{\|< x,y >\|} & = \left\langle \dfrac{x}{\|< x,y >\|}, \dfrac{y}{\|< x,y >\|} \right\rangle \\ OQ =< x',y' > & \Longrightarrow\ V = \dfrac{< x',y' >}{\|< x',y' >\|} & = \left\langle \dfrac{x}{\|< x',y' >\|}, \dfrac{y}{\|< x',y' >\|} \right\rangle \end{array}$$

The sines and cosines of these angles are found in terms of the four components of these two vectors U and V. These sines and cosines are exactly these components because the magnitude of the vector and, therefore, the hypotenuse is 1 in either case:

$$\begin{array}{ll} \cos\alpha = \dfrac{x}{\|< x,y >\|} & \sin\alpha = \dfrac{y}{\|< x,y >\|} \\ \cos\beta = \dfrac{x'}{\|< x',y' >\|} & \sin\beta = \dfrac{y'}{\|< x',y' >\|} \end{array}$$

Therefore, according to the formula, we have:

$$\begin{aligned} \cos\theta &= \frac{x}{\|< x,y >\|}\cdot\frac{x'}{\|< x',y' >\|} + \frac{y}{\|< x,y >\|}\cdot\frac{y'}{\|< x',y' >\|} \\ &= \frac{xx'+yy'}{\|< x,y >\|\cdot\|< x',y' >\|}\,. \end{aligned}$$

We will have a special name for the numerator of this fraction:

Definition 1.10.2: dot product

The *dot product* of vectors $< x, y >$ and $< x', y' >$ in $\mathbf{R}^2$ is defined by:

$$\boxed{< x, y > \cdot < x', y' >= xx' + yy'}$$

Thus, the dot product is computed, as other vector operations, componentwise.

We now re-state our theorem about the directions:

Theorem 1.10.3: Angles for Dimension 2

If θ is the angle between vectors $A \neq 0$ and $B \neq 0$ in $\mathbf{R}^2$, then:

$$\boxed{\cos\theta = \frac{A \cdot B}{||A||\ ||B||}}$$

Warning!

It makes sense not to put "$\cdot$" in the denominator to avoid confusion.

The presence of the magnitudes in the denominator suggests (to be proven later) that the result is, as expected, depends only on the directions.

Example 1.10.4: simple vectors

Let's test the theorem on simple vectors.

First the two basis vectors:

$$i =< 1, 0 >,\ j =< 0, 1 > \implies i \cdot j = 1 \cdot 0 + 0 \cdot 1 = 0\,.$$

Indeed, they are perpendicular and $\cos\pi/2 = 0$. Similarly,

$$< 1, 1 > \cdot < -1, 1 >= 1 \cdot (-1) + 1 \cdot 1 = 0\,.$$

However,

$$< 1, 0 > \cdot < 1, 1 >= 1 \cdot 1 + 0 \cdot 1 = 1\,.$$

To see the correct angle of 45 degrees, we apply the formula from the theorem:

$$\cos\theta = \frac{< 1, 0 > \cdot < 1, 1 >}{||< 1, 0 >||\ ||< 1, 1 >||} = \frac{1}{1\ \sqrt{2}} = \frac{\sqrt{2}}{2}\,.$$

The following is a very convenient result:

Corollary 1.10.5: Right Angle, Zero Dot Product

Two non-zero vectors are perpendicular if and only if their dot product is zero; i.e.,

$$\boxed{A \perp B \iff A \cdot B = 0}$$

Example 1.10.6: lines

Suppose we have two lines given by their relations:

$$y-3=2(x-1) \quad \text{and} \quad y+1=-3(x-3)\,.$$

What is the angle θ between them?

Do we need their parametric representations? No, just the direction vectors. The slope of the first is 2, so we can choose the direction vector to be $V=<1,2>$. The slope of the second is -3, so we can choose the direction vector to be $U=<1,-3>$. Therefore,

$$\cos\theta=\frac{<1,2>\cdot<1,-3>}{||<1,2>||\ ||<1,-3>||}=\frac{1-6}{\sqrt{5}\,\sqrt{10}}=-\frac{5}{\sqrt{50}}\,.$$

We start to climb the dimensions.

Definition 1.10.7: dot product

The *dot product* of vectors $<x,y,z>$ and $<x',y',z'>$ in $\mathbf{R}^3$ is defined by:

$$\boxed{<x,y,z>\cdot<x',y',z'>=xx'+yy'+zz'}$$

The dot product is componentwise operation:

$$\begin{array}{ccccccc} A & =< & x, & y, & z & > \\ \cdot \\ B & =< & u, & v, & w & > \\ \hline A\cdot B & = & x\cdot u+ & y\cdot v+ & z\cdot w \end{array} \qquad A\cdot B=\begin{bmatrix} x\\ y\\ z\end{bmatrix}\cdot\begin{bmatrix} u\\ v\\ w\end{bmatrix}=\begin{array}{l} x\cdot u+\\ y\cdot v+\\ z\cdot w\end{array}$$

Our theorem about the directions remains valid:

Theorem 1.10.8: Angles for Dimension 3

If θ is the angle between vectors $A\neq 0$ and $B\neq 0$ in $\mathbf{R}^3$, then:

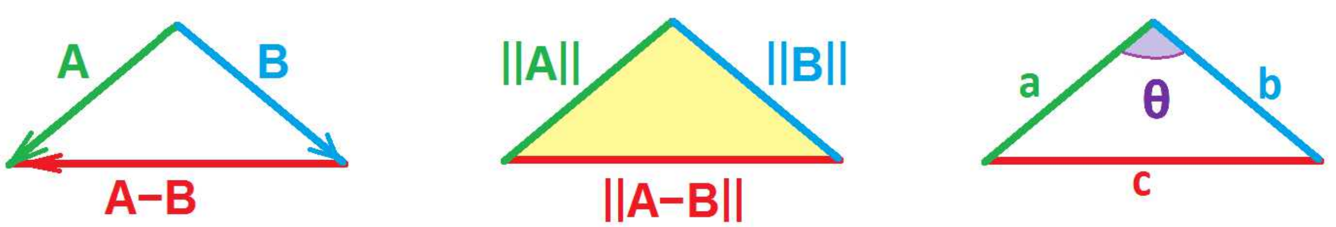
$$\boxed{\cos\theta=\frac{A\cdot B}{||A||\ ||B||}}$$

Proof.

Instead of trigonometric formulas we used for case $n=2$, we will rely on the *algebraic properties of the dot product*. We start with the *Law of Cosines* (cosine is what we are looking for anyway) which states:

$$c^2=a^2+b^2-2ab\cos\theta\,,$$

for any triangle with sides a,b,c and angle θ between a and b.



We interpret the lengths of the sides of the triangle in terms of the lengths of vectors:

$$a=||A||,\ b=||B||,\ c=||A-B||\,.$$

Then we translate the law into the language of vectors:

$$||A-B||^2 = ||A||^2 + ||B||^2 - 2||A||\ ||B|| \cos\theta\,.$$

Instead of solving for $\cos\gamma$, we expand the left-hand side:

$$\begin{array}{lll} ||A-B||^2 & = (A-B)\cdot(A-B) & \text{Normalization.}\\ & = A\cdot A + A\cdot(-B) + (-B)\cdot A + (-B)\cdot(-B) & \text{Distributivity.}\\ & = ||A||^2 - 2A\cdot B + ||B||^2 & \text{Associativity and Normalization.} \end{array}$$

The Law of Cosines then takes the following form:

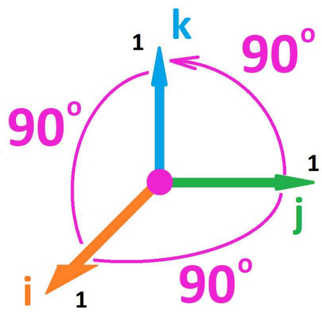
$$||A||^2 - 2A\cdot B + ||B||^2 = ||A||^2 + ||B||^2 - 2||A||\ ||B|| \cos\theta\,.$$

Now we cancel the repeated terms in the two sides of the equation and obtain the following:

$$-2A\cdot B = -2||A||\ ||B|| \cos\theta\,.$$

Example 1.10.9: basis vectors

It is once again easy to confirm that the basis vectors are perpendicular to each other:

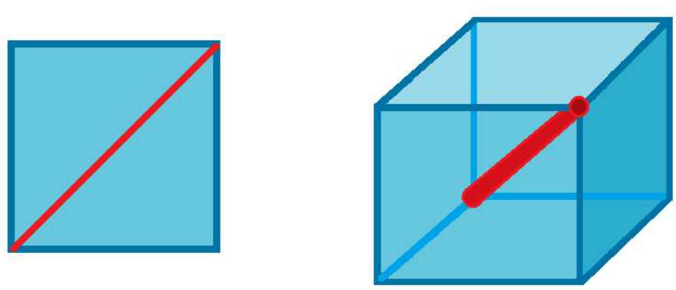


The 1 and 0 are mismatched:

$$\begin{array}{llll} i\cdot j & =<1,0,0>\cdot<0,1,0> & = 1\cdot 0 + 0\cdot 1 + 0\cdot 0 & = 0\\ j\cdot k & =<0,1,0>\cdot<0,0,1> & = 0\cdot 0 + 1\cdot 0 + 0\cdot 1 & = 0\\ k\cdot i & =<0,0,1>\cdot<1,0,0> & = 0\cdot 1 + 0\cdot 0 + 1\cdot 0 & = 0 \end{array}$$

Example 1.10.10: diagonals

The angle between the sides and the diagonal in a square is 45 degrees. Now, what is the angle between the diagonal of a *cube* and any of its edges? Try to guess from the picture if the angle is 45 degrees:



A hard trigonometric problem is solved easily with the dot-product.

First we choose vectors to represent the edges: the three basis vectors for the outside edges and $A=<1,1,1>$ for the diagonal. We have for the angle:

$$\cos\theta = \frac{<1,1,1>\cdot<1,0,0>}{||<1,1,1>||\ ||<1,0,0>||} = \frac{1}{\sqrt{3}}\,.$$

What is this angle? We only know that

$$\frac{1}{\sqrt{3}} < \frac{1}{\sqrt{2}}\,.$$

Because cosine is a decreasing function, it follows that this angle is larger than $\pi/4$. There is more room for maneuver than on the plane!

Exercise 1.10.11

Find all the angles from the center of a cube to its corners.

Can we make sense of directions and angles in $\mathbf{R}^n$?

We previously "extrapolated" the definition of the magnitude (and the distance before that) to produce the Euclidean norm:

dimension	vector	components	norm
1	A	a	$\|A\| = \|a\|$
2	A	$< a, b >$	$\|\|A\|\|^2 = a^2 + b^2$
3	A	$< a, b, c >$	$\|\|A\|\|^2 = a^2 + b^2 + c^2$
...	...	...	...
n	A	$< a_1, a_2, ..., a_n >$	$\|\|A\|\|^2 = a_1^2 + a_2^2 + ... + a_n^2$

We do the same for the definition of the dot product:

dimension	vectors	components	dot product
1	A	a	
	B	u	$A \cdot B = a \cdot u$
2	A	$< a, b >$	
	B	$< u, v >$	$A \cdot B = a \cdot u + b \cdot v$
3	A	$< a, b, c >$	
	B	$< u, v, w >$	$A \cdot B = a \cdot u + b \cdot v + c \cdot w$
...	...	...	...
n	A	$< a_1, a_2, ..., a_n >$	
	B	$< b_1, b_2, ..., b_n >$	$A \cdot B = a_1 \cdot b_1 + a_2 \cdot b_2 + ... + a_n \cdot b_n$

Definition 1.10.12: dot product in n-space

The *dot product* of vectors A and B is defined to be the sum of the products of

their components:

$$\begin{array}{rcllll}
A & =< & a_1, & a_2, & ..., & a_n & > \\
B & =< & b_1, & b_2, & ..., & b_n & > \\
A \cdot B & = & a_1b_1 & +a_2b_2 & +... & +a_nb_n &
\end{array}$$

In the sigma notation:

$$A \cdot B = \sum_{k=1}^{n} a_k b_k$$

Exercise 1.10.13

What is the angle between the diagonal of a 4-dimensional cube and any of its edges?

Exercise 1.10.14

What is the angle between the diagonal of a n-dimensional cube and any of its edges? What value does the angle approach when n approaches infinity?

Below we see how this new operation (last row) compares with the other vector operations:

vector addition	A	$+$	B	$=$	C
	vector		vector		vector
scalar multiplication	c	$\cdot$	A	$=$	C
	number		vector		vector
dot product	A	$\cdot$	B	$=$	s
	vector		vector		number

Warning!

The last two might be confusing without a *context*; for example, consider the three possible meanings of the following:

$$0 \cdot A = 0\,.$$

Let's consider the *properties of the dot product.*

If we just set $Y = X$, we have the so-called *Normalization*:

$$||X||^2 = X \cdot X$$

One can, therefore, recover the magnitude from the dot product just as before. So, once we have the dot product, we don't need to introduce the magnitude independently.

The *Positivity* of the norm then requires a similar property for the dot product:

$$V \cdot V \geq 0; \quad \text{and} \quad V \cdot V = 0 \iff V = 0$$

Next, *Commutativity* or *Symmetry*:

$$\boxed{A \cdot B = B \cdot A}$$

This means that the angle is *between* A and B; i.e., the same from A to B as from B to A.

Next, *Associativity*:

$$\boxed{(kA) \cdot B = k(A \cdot B) = A \cdot (kB)}$$

So, the effect of stretching on the dot product is a multiple and the angle doesn't change for $k > 0$ or is replaced with the opposite when $k < 0$.

We can see now that only the normalizations matter for the angle between two vectors. We just choose these values for k in the last formula:

$$\frac{1}{||A||\ ||B||},\ \frac{1}{||A||},\quad \text{and}\quad \frac{1}{||B||}$$

to rewrite our formula from the last section:

$$\cos\theta = \frac{A \cdot B}{||A||\ ||B||} = \frac{1}{||A||\ ||B||} A \cdot B = \frac{1}{||A||} A \cdot \frac{1}{||B||} B = \frac{A}{||A||} \cdot \frac{B}{||B||}\,.$$

The result suggests that *the dot product is independent from the coordinate system.* Certainly, this system is just a tool that we introduce into the space the geometry of which we study, and we don't expect that changing the components of vectors will also change the distances and the angles. But it's also true in $\mathbf{R}^n$!

Next, *Distributivity* or *Linearity*:

$$\boxed{A \cdot (B + C) = A \cdot B + A \cdot C}$$

Treated componentwise, the Commutativity, Associativity, Distributivity properties for the dot product of vectors follow from the Commutativity, Associativity, Distributivity for numbers. For example, this is the whole proof of the Commutativity for $n = 2$:

$$\begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = au + bv = ua + vb = \begin{bmatrix} u \\ v \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix}$$

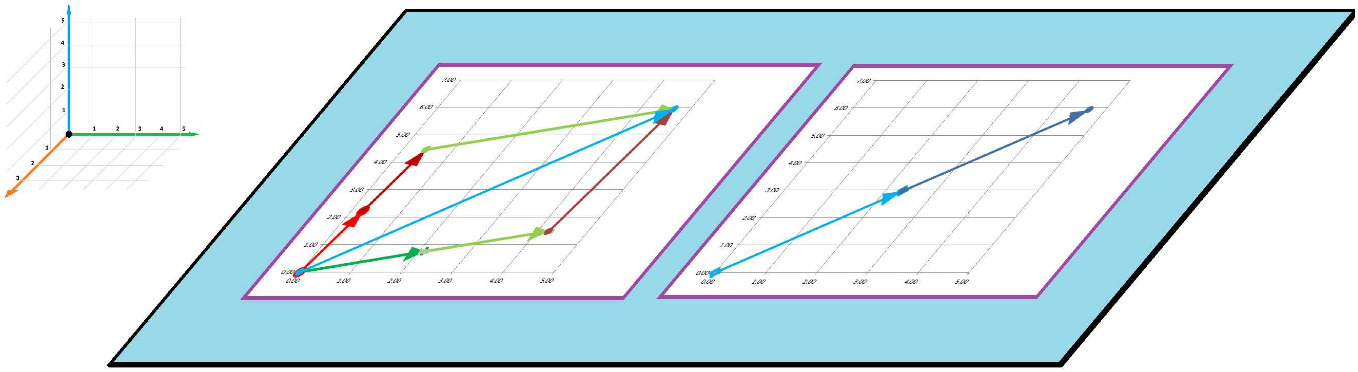
Once again, these properties allow us to use the usual algebraic manipulation steps for numbers as long as the expressions make sense to begin with.

How do we understand the geometry of this made-up space? After all, when $n > 3$, there is no reality test for this concept and we can't verify the formulas we are to use!

We have come to understand the distances in $\mathbf{R}^n$ and now ask:

- ▶ What is the meaning of the angle between two vectors A and B in $\mathbf{R}^n$?

The answer is to reduce the multidimensional case to the case of $n = 2$. Indeed, every two vectors define a plane and this plane has the same vector algebra operations – including the dot product – as the ambient space $\mathbf{R}^n$:



The Distributivity will require 3 dimensions. In the meantime, the plane has the well-understood Euclidean geometry: The lengths of vectors and the angles between vectors can be measured.

The definition is abstract but it matches the lower dimensions $n = 1, 2, 3$:

Definition 1.10.15: angles for dimension n

The angle θ between vectors $A \neq 0$ and $B \neq 0$ in $\mathbf{R}^n$ is defined to satisfy:

$$\boxed{\cos\theta = \frac{A \cdot B}{||A||\ ||B||}}$$

For the record, we summarize the rules of the dot product:

Theorem 1.10.16: Axioms of Inner Product Space

The dot product in $\mathbf{R}^n$ satisfies the following three properties for all vectors U, V, W and all scalars a, b:

- ***Symmetry:*** $U \cdot V = V \cdot U$.
- ***Linearity:*** $U \cdot (aV + bW) = a(U \cdot V) + b(U \cdot W)$.
- ***Positive-definiteness:*** $V \cdot V \geq 0$; *and* $V \cdot V = 0 \iff V = 0$.

We have added a third vector operation to the toolkit but vector algebra still looks like that of numbers!

From the inequality

$$|\cos\theta| \leq 1\,,$$

we derive the following.

Corollary 1.10.17: Cauchy Inequality

For any pair of vectors $A \neq 0$ and $B \neq 0$ in $\mathbf{R}^n$, we have:

$$\boxed{|A \cdot B| \leq ||A||\ ||B||}$$

In other words, if we rotate one of the vectors, the dot product reaches its maximum when they are parallel to each other.

So, what *is* the dot product of two vectors? Two partial answers:

1. The dot product is the cosine of the angle when the vectors are unit vectors.
2. The dot product is the square of the magnitude when the vectors are equal.

A complete, geometric answer may simply be our formula:

$$\boxed{A \cdot B = ||A||\ ||B|| \cos\theta}$$

Example 1.10.18: inner product for taxicab?

Once can see how the dot product emerged from the Euclidean metric and norm:

dimension	vectors	components	Euclidean norm and dot product
n	A	$< a_1, a_2, ..., a_n >$	$\|\|A\|\|^2 = a_1 \cdot a_1 + a_2 \cdot a_2 + ... + a_n \cdot a_n$
n	A	$< a_1, a_2, ..., a_n >$	
	B	$< b_1, b_2, ..., b_n >$	$A \cdot B = a_1 \cdot b_1 + a_2 \cdot b_2 + ... + a_n \cdot b_n$

But what about the taxicab metric? We can suggest the following candidate for the inner product:

dimension	vectors	components	taxicab norm and inner product?
n	A	$< a_1, a_2, ..., a_n >$	$\|\|A\|\| = \|a_1\| + \|a_2\| + ... + \|a_n\|$
n	A	$< a_1, a_2, ..., a_n >$	
	B	$< b_1, b_2, ..., b_n >$	$A \cdot B = \sqrt{\|a_1\| \cdot \|b_1\|} + \sqrt{\|a_2\| \cdot \|b_2\|} + ... + \sqrt{\|a_n\| \cdot \|b_n\|}$

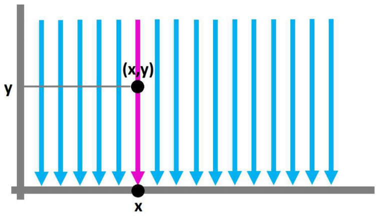
Unfortunately, the Linearity fails! We, therefore, will be unable to discuss angles in this space.

Exercise 1.10.19

Prove the last statement for $n = 2$.

1.11. *Projections and decompositions of vectors*

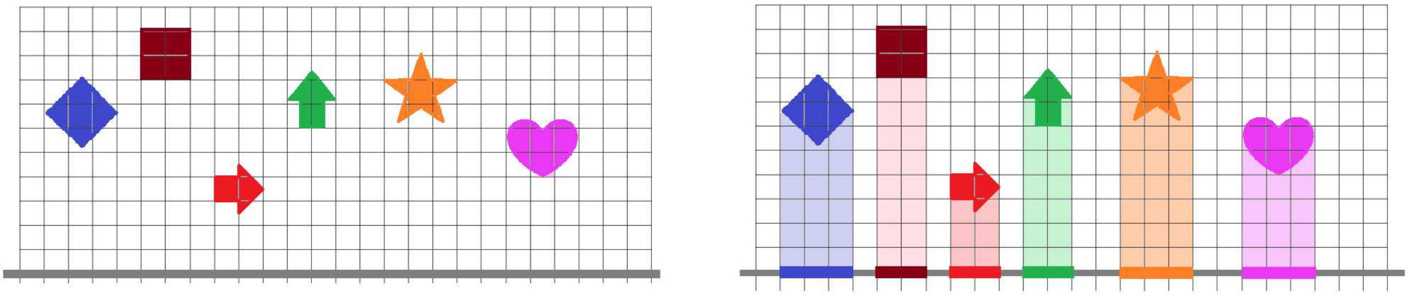
To find the x-coordinate of a point on the xy-plane, we go vertically from that points until we reach the x-axis:



It's a familiar function:

$$F : \mathbf{R}^2 \to \mathbf{R},\ < x, y > \mapsto x\,.$$

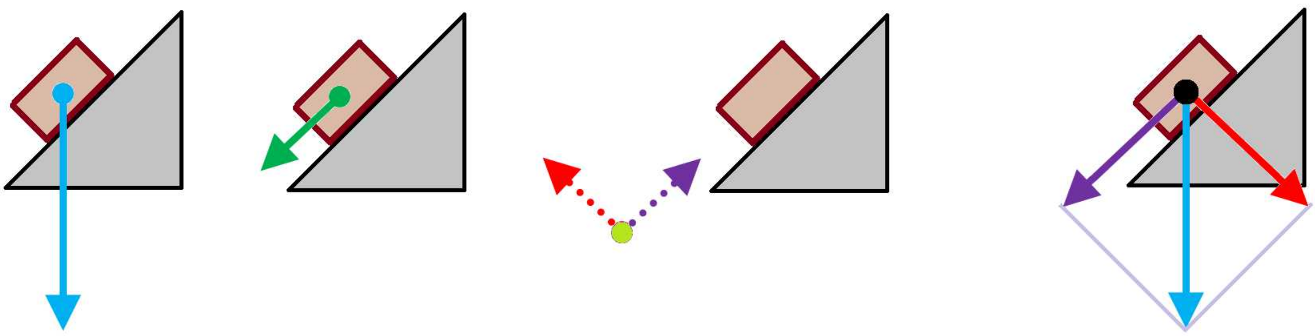
The result resembles *shadows* cast by points, vectors, or subsets onto the x-axis with the light cast from above (or from the right for the y-axis):



It is called the *projection* of the point on the x-axis. Same for the y-axis.

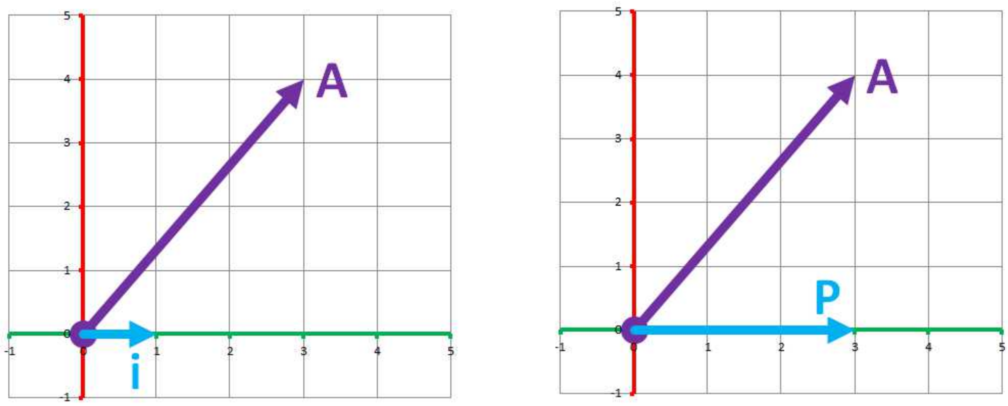
Similarly, the projection of a *vector* on the x-axis gives its x-component. If several coordinate systems co-exist, a transition from one to another will require expressing the new coordinates of a point or the new components of a vector in terms of the old. We do that one axis or basis vector at a time.

As a reason for this study here is the example of compound motion from earlier in this chapter, an object sliding down a slope:



In order to concentrate on the relevant part of the motion, one would choose the first basis vector to be parallel to the surface and the second perpendicular.

For vectors, the projection on the x-axis will require choosing a representative vector on it; it's i. Any vector A is expressed in terms of i by finding A's component P on the x-axis. For example, below the component is 3:

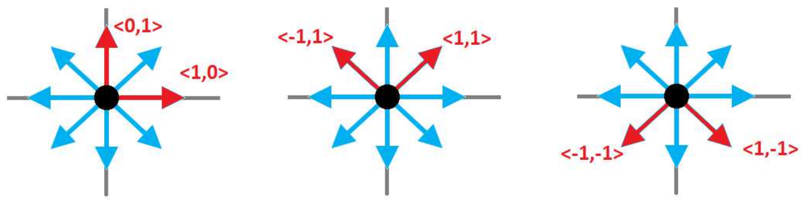


Specifically, we have a linear combination of i and j:

$$A = < 3, 4 > = 3i + 4j .$$

What if instead of i we have an arbitrary vector V?

But first, we consider a shortcut for finding a *perpendicular vector*:



The problem of rotating a given vector V in the plane through $\pi/2$ has an easy solution:

Theorem 1.11.1: Orthogonality in 2-space

For any vector $V =< u, v >$ on the plane, the following vector, called a normal vector of V, denoted as follows $V^\perp$, has the angle of $\pi/2$ with V:

$$\boxed{V^\perp =< u, v >^\perp =< -v, u >}$$

Proof.

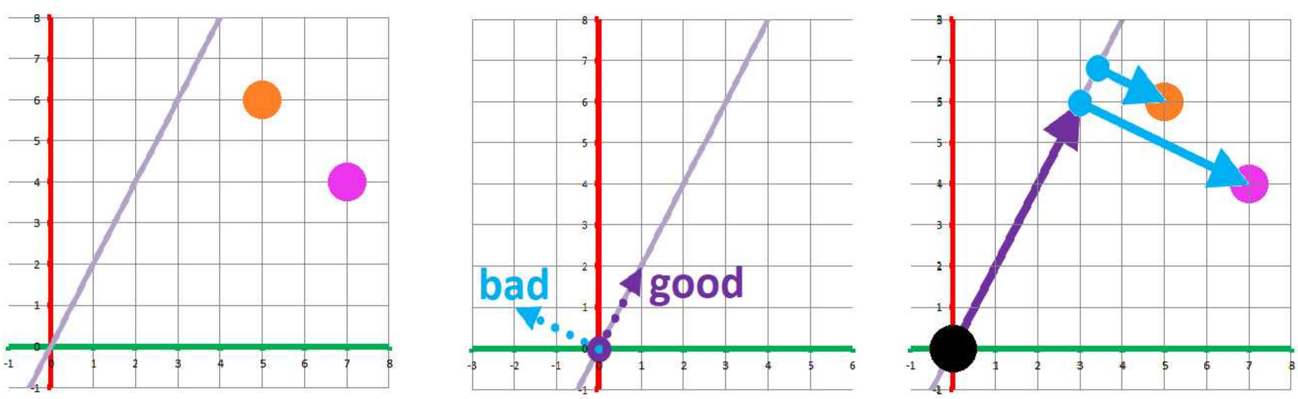
This is easy to confirm:

$$V \cdot V^\perp =< u, v > \cdot < -v, u >= u(-v) + vu = 0\,.$$

We have then a special operation on vectors in dimension 2. Every vector has exactly two normal *unit* vectors.

Example 1.11.2: investing advice

An initially investment advice might be very simple, for example: hold the proportion of stocks and bonds 1-to-2. If x is the amount of stocks and y is the amount of bonds in it, the "ideal" portfolios lie on the line $y = 2x$, i.e., they are the ends of vectors that are multiples of $V =< 2, 1 >$:



How do we determine how close is each portfolio to the ideal? We can use V as a measuring stick. What about the bad? We can find a vector perpendicular to V via the last theorem. Though not unit vectors, these two vectors will give the good and the bad components of any portfolio:

$$g =< 1, 2 > \quad \text{and} \quad b =< -2, 1 > \ .$$

So, we need to represent every portfolio as a linear combination of these two. For example:

$$< 5, 6 >= p < 1, 2 > +q < -2, 1 > \quad \text{and} \quad < 7, 4 >= u < 1, 2 > +v < -2, 1 > \ .$$

In other words, we need to solve two systems of linear equations:

$$\begin{cases} p - 2q = 5 \\ 2p + q = 6 \end{cases} \quad \text{and} \quad \begin{cases} u - 2v = 7 \\ 2u + v = 4 \end{cases}$$

These are the solutions:

$$p = 17/5,\ q = -4/5 \quad \text{and} \quad u = 4,\ v = -2\,.$$

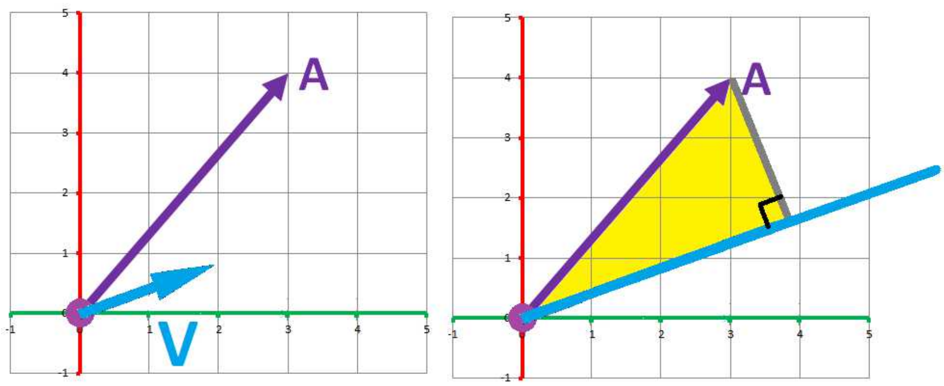
The second numbers in these pairs indicate how far it is from the ideal. The second portfolio is farther.

What if we have many investment vehicles in each portfolio? For example, we may assume that these portfolios live in $\mathbf{R}^7$:

					$	$	$
		%	weights	unit	25	Portfolios	
		V	V/100	V/||V||	line	A1	A2
	magnitude	41.23	0.41	1.00	10.31	10.39	5.20
1	growth stocks	20	0.20	0.49	5	4	0
2	value stocks	10	0.10	0.24	3	2	3
3	corporate bonds	5	0.05	0.12	1	4	2
4	minicipal bonds	15	0.15	0.36	4	2	1
5	federal bonds	25	0.25	0.61	6	8	0
6	real estate	15	0.15	0.36	4	0	3
7	cash	10	0.10	0.24	3	2	2
	SUM	100	1.00	2.43	25	22	11

Unfortunately, the shortcut of the last theorem is not available anymore. We will need a further analysis.

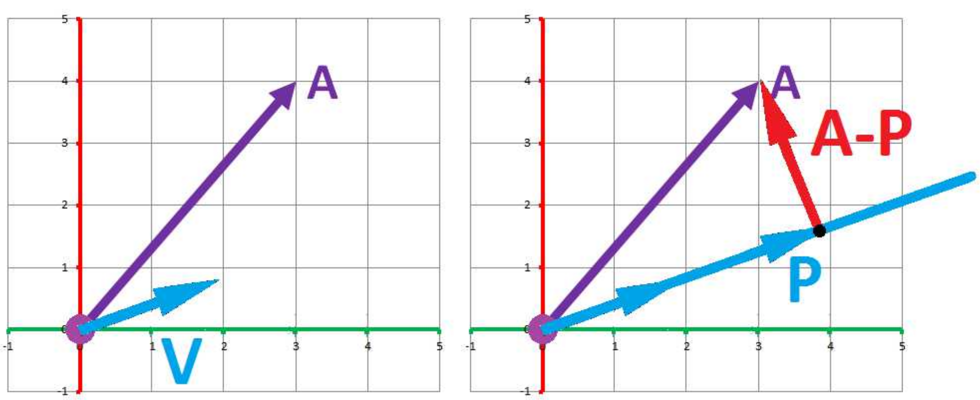
This is what a projection on a line defined by a vector looks like:



The question becomes: How much does A "protrude" in the direction of V? More precisely:

- ▶ Vector A is expressed in terms of V by finding A's projection P on the line created by V.

So, every vector A needs to be expressed as a linear combination of V and some other vector perpendicular to V:



If P is the projection, what's the other vector? We simply subtract:

$$A = P + (A - P)\,.$$

The vector we are after is described indirectly:

Definition 1.11.3: orthogonal projection

Suppose A and $V \neq 0$ are two vectors in $\mathbf{R}^n$. Then the *orthogonal projection of A onto V* is a vector P that satisfies the following:

1. P is parallel to V.
2. $A - P$ is perpendicular to V.

Let's find an explicit formula.

First, "parallel" simply means a multiple! Therefore, the first property means that there is a number k – this is the one we are looking for – such that:

$$P = kV\,.$$

The second property is expressed in terms of the dot product:

$$V \cdot (A - P) = 0\,.$$

We substitute:

$$V \cdot (A - kV) = 0\,,$$

and use *Distributivity* and *Associativity*:

$$V \cdot A - kV \cdot V = 0\,.$$

Next we use *Normalization*:

$$V \cdot A = k||V||^2\,.$$

Then,

$$k = \frac{V \cdot A}{||V||^2}\,.$$

This is the multiple of V that gives us P. Thus, we have proven the following result:

Theorem 1.11.4: Projection Via Dot Product

The orthogonal projection of a vector A onto a vector $V \neq 0$ is the vector P given by:

$$\boxed{P = \frac{V \cdot A}{||V||^2} V}$$

Notice that the formula – as expected – depends only on the *direction* of V; only unit vectors are involved:

$$P = \left(\frac{V}{||V||} \cdot A \right) \frac{V}{||V||}\,.$$

Exercise 1.11.5

Prove the last formula.

Example 1.11.6: investing advice continued

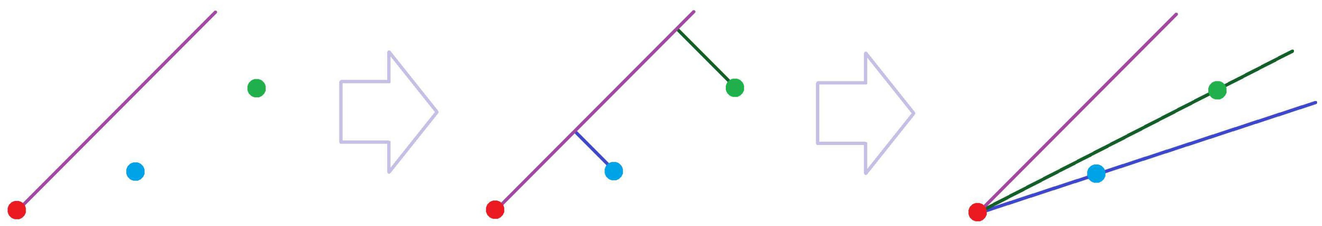
We have 7 instruments in each portfolio. The portfolio *advice* V is shown in column 4 and the two competing portfolios A_1 and A_2 in columns 8 and 9. We carry out the following computations:

1. We find the coordinatewise products of A_1 and A_2 with V (columns 10 and 11) and then by adding those obtain the two dot products (bottoms of the columns).

2. From those two, we find the multiples c_1 and c_2 for the projections according to the last theorem (tops of columns 12 and 13).
3. We use those two to multiply V componentwise to obtain the projections P_1 and P_2 of A_1 and A_2 (columns 12 and 13).
4. The differences of A_1 and A_2 from V are found (columns 14 and 15).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2											c1	c2		
3						$	$	$	Dot products		0.218	0.071		
4			%	weights	unit	25	Portfolios		and angles		Projections		Distances	
5			V	V/100	V/||V||	line	A1	A2	V.A1	V.A2	P1	P2	D1	D2
6		magnitude	41.23	0.41	1.00	10.31	10.39	5.20	0.86	0.56			5.24	4.30
7	1	growth stocks	20	0.20	0.49	5	4	0	80	0	4.35	1.41	0.35	1.41
8	2	value stocks	10	0.10	0.24	3	2	3	20	30	2.18	0.71	0.18	-2.29
9	3	corporate bonds	5	0.05	0.12	1	4	2	20	10	1.09	0.35	-2.91	-1.65
10	4	minicipal bonds	15	0.15	0.36	4	2	1	30	15	3.26	1.06	1.26	0.06
11	5	federal bonds	25	0.25	0.61	6	8	0	200	0	5.44	1.76	-2.56	1.76
12	6	real estate	15	0.15	0.36	4	0	3	0	45	3.26	1.06	3.26	-1.94
13	7	cash	10	0.10	0.24	3	2	2	20	20	2.18	0.71	0.18	-1.29
14		SUM	100	1.00	2.43	25	22	11	370	120	21.76	7.06		

Finally, we compute the magnitudes of those two (tops of columns 14 and 15). We conclude that the *second* portfolio is closer to the line that represents the ideal:



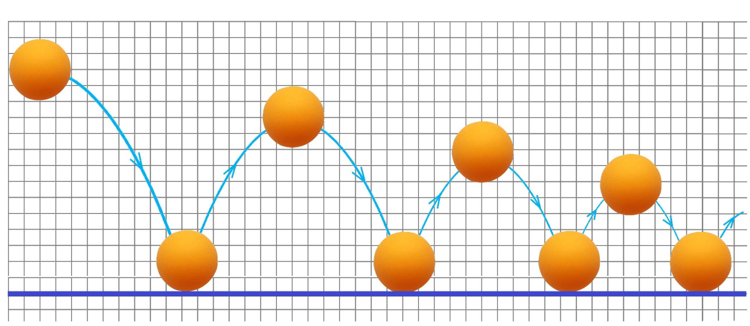
Alternatively, we realize that these distances are proportional to the size of the investment, which skews the conclusions. We then turn to the *angles* instead. Their cosines are computed (tops of columns 10 and 11) with the formula from the last section. We conclude that the *first* portfolio is closer to the line that represents the ideal.

Exercise 1.11.7

What is the set of all portfolios with the total investment of 1?

1.12. *Sequences and topology in* $\mathbf{R}^n$

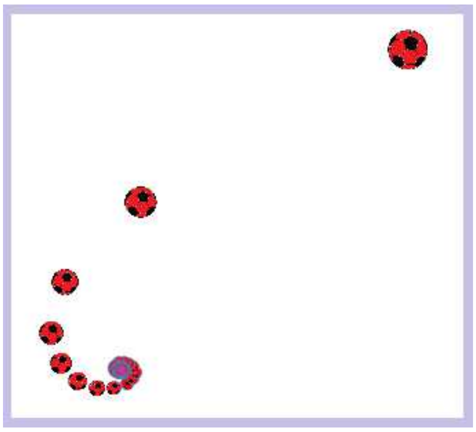
Recall from Volume 1 (Chapter 1PC-1) the image of a ping-pong ball bouncing off the floor. Recording its height every time gives us a *sequence*:



It is a *sequence of numbers*!

Imagine now watching a ball bouncing on an uneven surface with its locations recorded at equal periods of

time. The result is a sequence of locations on the plane:



We will study infinite sequences of locations and especially their *trends*. An *infinite* sequence will be sometimes "accumulating" around a single location. The gap between the ball and the drain becomes invisible!

Definition 1.12.1: infinite sequence

A function defined on a ray in the set of integers, $\{p, p+1, ...\}$, is called an *infinite sequence*, or simply sequence with the notation:

$$A_k: \ k = p, p+1, p+2, \ldots,$$

or, abbreviated,

$$A_k\,.$$

Every function $X = f(t)$ with an appropriate domain creates a sequence:

$$A_k = f(k)\,.$$

We visualize numerical sequences as *sequences of points* on the xy-plane via their graphs.

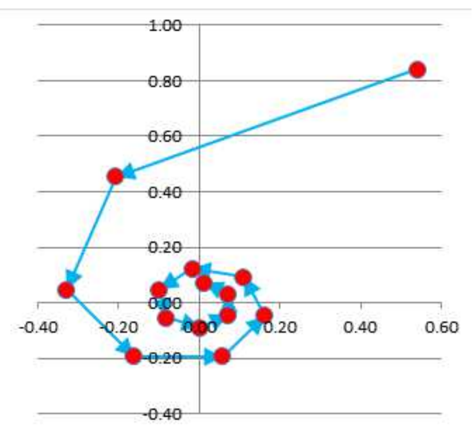
Example 1.12.2: reciprocals

The go-to example is that of the sequence made of the reciprocals:

$$A_k = \left(\frac{\cos k}{k}, \ \frac{\sin k}{k} \right)\,.$$

It *tends to* 0 while spiraling around it.

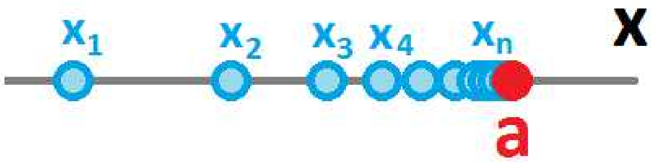
n	x	y
1.00	0.54	0.84
2.00	-0.21	0.45
3.00	-0.33	0.05
4.00	-0.16	-0.19
5.00	0.06	-0.19
6.00	0.16	-0.05
7.00	0.11	0.09
8.00	-0.02	0.12
9.00	-0.10	0.05
10.00	-0.08	-0.05
11.00	0.00	-0.09
12.00	0.07	-0.04
13.00	0.07	0.03
14.00	0.01	0.07



This fact is easily confirmed numerically.

Unfortunately, not all sequences are as simple as that. They may approach their respective limits in an infinite variety of ways. And then there are sequences with no limits. We need a more general approach.

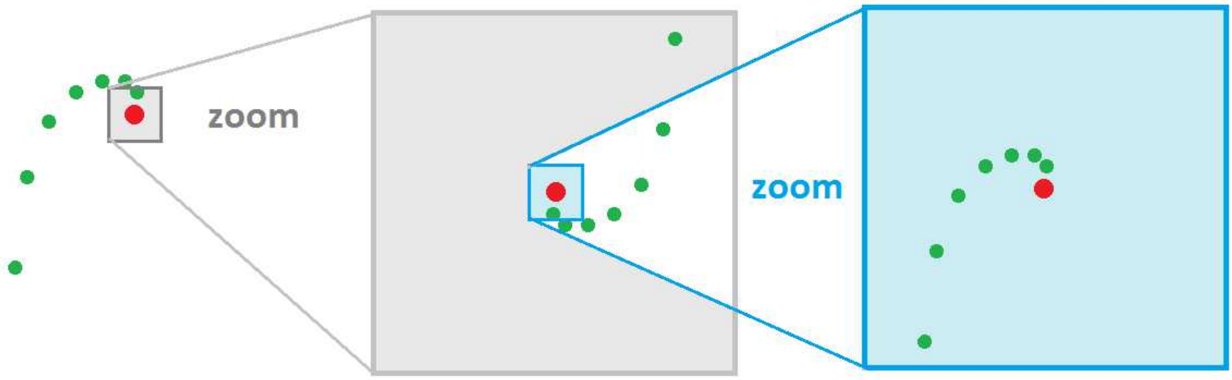
Let's recall from Volume 2 (Chapter 2DC-1) the idea of the convergence of infinite sequences of real numbers:



The idea is that the distance from the kth point to the limit is getting smaller and smaller:

$$|x_k - a| \to 0 \ \text{ as } k \to \infty\,.$$

In the n-dimensional case, the idea remains the same: a sequence of points in $\mathbf{R}^n$ is getting closer and closer to its limit, which is also a point in $\mathbf{R}^n$:



We rewrite what we want to say about the meaning of the limits in progressively more and more precise terms.

k	$X = A_k$
As $k \to \infty$,	we have $X \to A$.
As k approaches ∞,	X approaches A.
As k is getting larger and larger,	the distance from X to A approaches 0.
By making k larger and larger,	we make $d(X, A)$ as small as needed.
By making k larger than some $N > 0$,	we make $d(X, A)$ smaller than any given $\varepsilon > 0$.

Then, the following condition holds:

▶ for each real number $\varepsilon > 0$, there exists a number N such that, for every natural number $k > N$, we have

$$d(A_k, A) < \varepsilon\,.$$

Our study become much easier once we realize that distances are numbers and $d(A_k, A)$ is just a sequence of numbers! Understanding distances in $\mathbf{R}^n$ and numerical sequences allows us easily to sort this out.

Definition 1.12.3: convergent sequence and limit

Suppose $A_k\ :\ \ k = 1, 2, 3...$ is a sequence of points in $\mathbf{R}^n$. We say that the sequence *converges* to another point A in $\mathbf{R}^n$, called the *limit* of the sequence, if:

$$d(A_n, A) \to 0 \ \text{ as } k \to \infty\,,$$

denoted as follows:

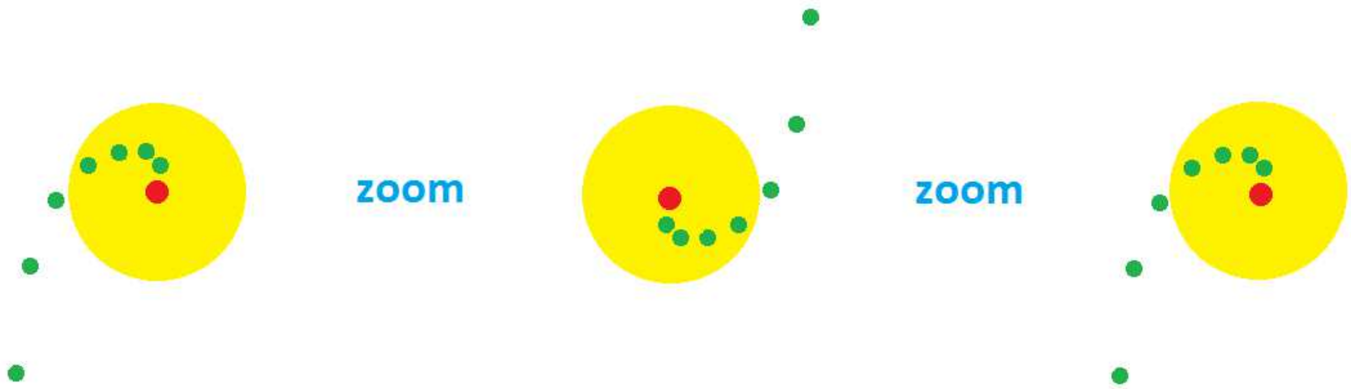
$$A_k \to A \ \text{ as } k \to \infty\,,$$

or

$$A = \lim_{k\to\infty} A_k\,.$$

If a sequence has a limit, then we call the sequence *convergent* and say that it *converges*; otherwise it is *divergent* and we say it *diverges*.

In other words, the points start to accumulate in smaller and smaller circles around A. A way to visualize a trend in a convergent sequence is to enclose the tail of the sequence in a *disk*:

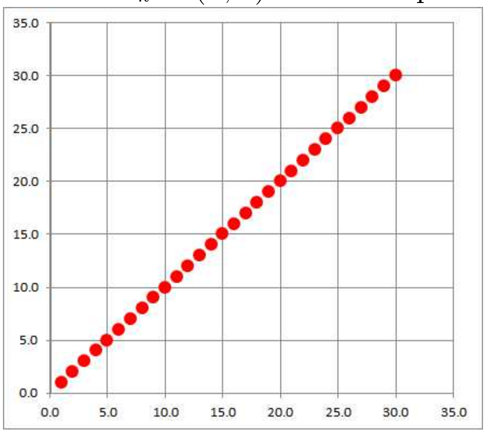


It should be, in fact, a narrower and narrower band; its width is 2ε. Meanwhile, the starting point of the band moves to the right; that's N.

Examples of divergence are below.

Example 1.12.4: go to infinity

A sequence may tend to infinity, such as $A_k = (k,k)$ at the simplest:



Then no disk – no matter how large – will contain the sequence's tail.

This behavior however has a meaningful pattern.

Definition 1.12.5: sequence tends to infinity

We say that a sequence A_n *tends to infinity* if the following condition holds:

- ▶ for each real number R, there exists a natural number N such that, for every natural number $k > N$, we have

$$d(0, A_n) > R\,.$$

We describe such a behavior with the following notation:

Limit

$$A_k \to \infty \quad \text{as } k \to \infty$$

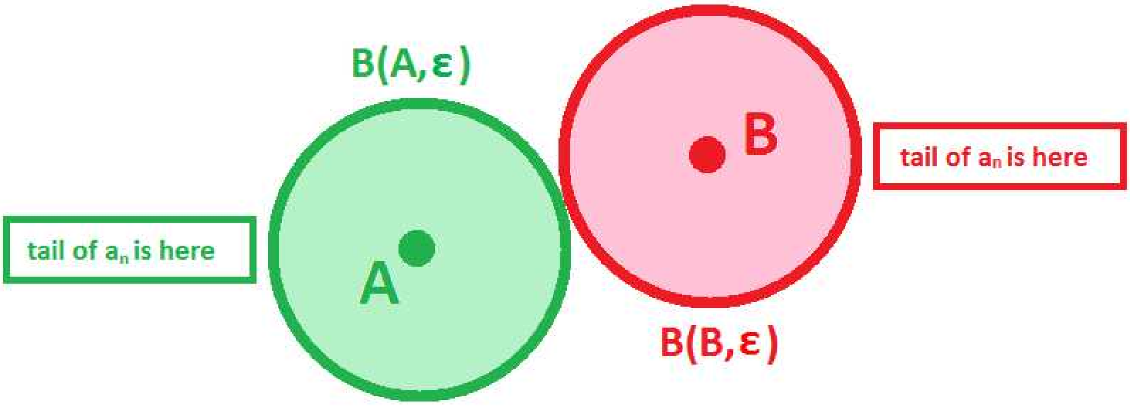
We need to justify "the" in "the limit".

Theorem 1.12.6: Uniqueness of Limit of Sequence

A sequence can have only one limit (finite or infinite); i.e., if A and B are limits of the same sequence, then $A = B$.

Proof.

The geometry of the proof is the following: we want to separate these two points by two non-overlapping disks. Then the tail of the sequence would have to fit one or the other, but not both. In order for them to be disjoint, their radii (that's ε!) should be less than half the distance between the two points.



The proof is by contradiction. Suppose A and B are two limits, i.e., either satisfies the definition. Let

$$\varepsilon = \frac{d(A,B)}{2}\,.$$

Now, we write the definition twice:

- there exists a number L such that, for every natural number $k > L$, we have

$$d(A_k, A) < \varepsilon\,,$$

  and
- there exists a number M such that, for every natural number $k > M$, we have

$$d(A_k, B) < \varepsilon\,.$$

In order to combine the two statements we need them to be satisfied for the same values of k. Let

$$N = \min\{L, M\}\,.$$

Then, for every number $k > N$, we have

$$d(A_k, A) < \varepsilon \quad \text{and} \quad d(A_k, B) < \varepsilon\,.$$

In particular, for every $k > N$, we have by the *Triangle Inequality*:

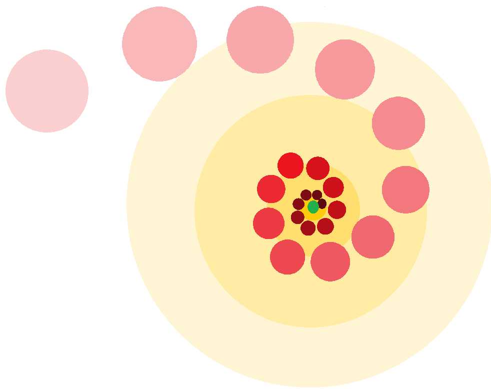
$$d(A,B) \le d(A, A_k) + d(A_k, B) < \varepsilon + \varepsilon < 2\varepsilon\,.$$

A contradiction.

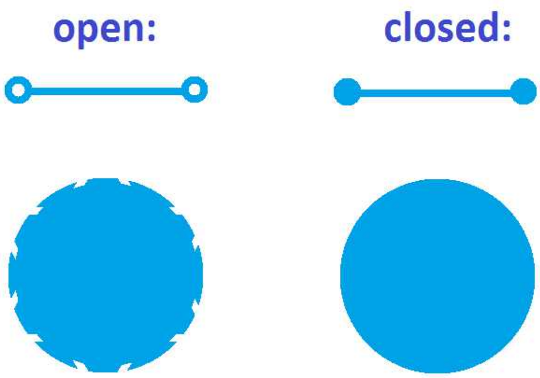
Exercise 1.12.7

Follow the proof and demonstrate that that it is impossible to for a sequence to have as limit: a point and infinity.

Thus, there can be no two limits and we are justified to speak of *the* limit.



What is the n-dimensional analog of a closed interval?



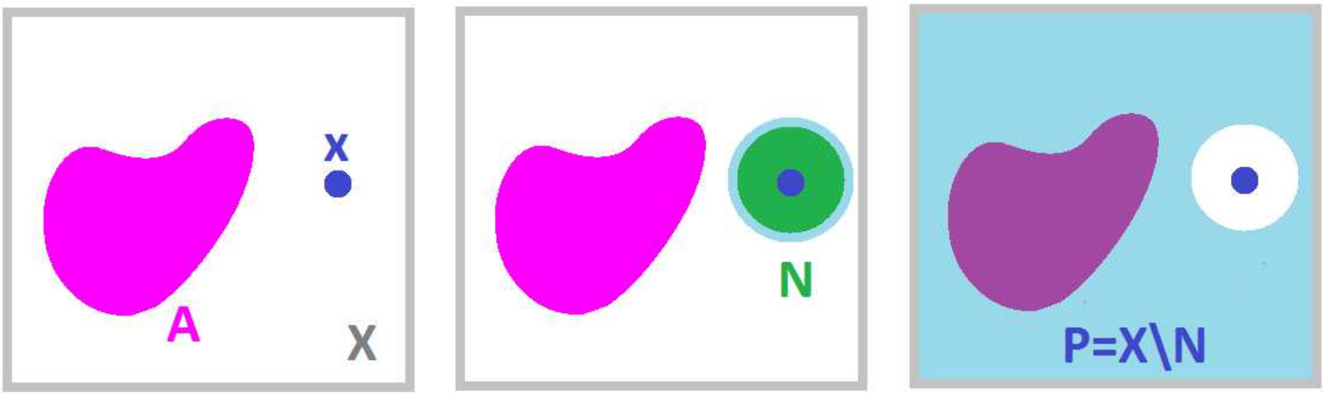
Compare the *disk* and the disk with its boundary (the circle) removed:

$$\{(x,y):\ x^2+y^2 \leq 1\} \ \text{ vs. } \ \{(x,y):\ x^2+y^2 < 1\}\,.$$

Or think of the ball and the ball with its boundary (the sphere) removed. The difference is that in the latter one can reach the boundary – and the outside of the set – by following a sequence that lies entirely inside the set!

Definition 1.12.8: closed set

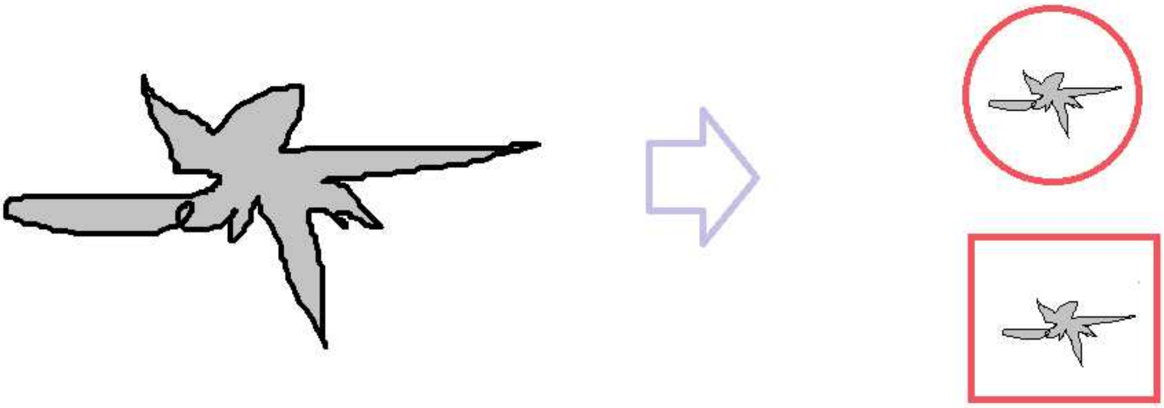
A set in $\mathbf{R}^n$ is called *closed* if it contains the limits of all of its convergent sequences.



Definition 1.12.9: bounded set

A set S in $\mathbf{R}^n$ is *bounded* if it fits in a sphere (or a box) of a large enough size:

$$d(x,0) < Q \quad \text{for all } x \text{ in } S\,.$$



1.13. The coordinatewise treatment of sequences

Here we go *back* from the treatment of the space to the spread-out coordinatewise approach.

Suppose space $\mathbf{R}^n$ is supplied with a Cartesian system.

Let's first look at the definition of the limit of a sequence X_k in $\mathbf{R}^n$. The limit is defined to be such a point A in $\mathbf{R}^n$ that

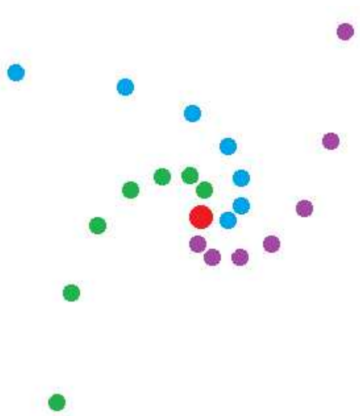
$$d(X_k, A) \to 0 \quad \text{as } k \to \infty\,.$$

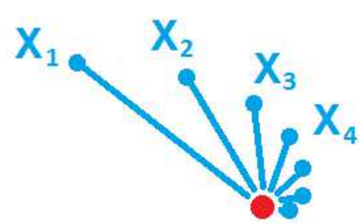
Suppose we use the Euclidean metrics is our space $\mathbf{R}^3$, what does the above condition mean? Suppose

$$X_k = \big(x_k, y_k, z_k\big) \quad \text{and} \quad A = (a, b, c)\,.$$

Then we have:

$$\sqrt{(x_k - a)^2 + (y_k - b)^2 + (z_k - c)^2} \to 0\,.$$





This limit is equivalent to the following:

$$(x_k - a)^2 + (y_k - b)^2 + (z_k - c)^2 \to 0\,,$$

because the function u^2 is continuous at 0. Since these three terms all non-negative, all three have to approach 0! Then we have:

$$(x_k - a)^2 \to 0,\ (y_k - b)^2 \to 0,\ (z_k - c)^2 \to 0\,.$$

These limits are equivalent to the following:

$$|x_k - a| \to 0,\ |y_k - b| \to 0,\ |z_k - c| \to 0\,,$$

because the function $\sqrt{u}$ is continuous from the right at 0. Finally, we have:

$$x_k \to a,\ y_k \to b,\ z_k \to c\,.$$

All coordinate sequences converge!

For the n-dimensional case, build a table:

	1	2	...	n	
X_k	x_k^1	x_k^2	...	x_k^n	
$\downarrow$	$\downarrow$	$\downarrow$	...	$\downarrow$	as $k \to \infty$
A	a^1	a^2	...	a^n	

The following is a summary.

Theorem 1.13.1: Coordinatewise Convergence

A sequence of points X_k in $\mathbf{R}^n$ converge to another point A if and only if every coordinate of X_k converges to the corresponding coordinate of A; i.e.,

$$X_k \to A \quad \text{as } k \to \infty \iff x_k^i \to a^i \quad \text{as } k \to \infty \quad \text{for all } i = 1, 2, ..., n\,,$$

where

$$X_k = (x_k^1,\ x_k^2,\ ...,\ x_k^n) \quad \text{and} \quad A = (a^1,\ a^2,\ ...,\ a^n)\,.$$

Exercise 1.13.2

Prove the "if" part of the theorem.

Example 1.13.3: a computation

We compute using the *Continuity Rule for Numerical Sequences*:

$$\begin{aligned}\lim_{k\to\infty}\left(\cos\frac{1}{k},\ \sin\frac{1}{k}\right) &= \left(\lim_{k\to\infty}\cos\frac{1}{k},\ \lim_{t\to\infty}\sin\frac{1}{k}\right)\\ &= (\cos 0,\ \sin 0)\\ &= (1, 0),\end{aligned}$$

because $\sin t$ and $\cos t$ are continuous.

This theorem also makes it easy to prove the algebraic properties of limits from those for numerical functions.

Now, sequences of vectors and their limits.

Vectors correspond to points:

$$OP \longleftrightarrow P$$

We are then able to discuss *convergence of sequences of vectors*. For convenience, we just restate the definition given earlier in this chapter replacing distances between points with magnitudes of differences of vectors:

$$d(PQ) = ||OQ - OP||\,.$$

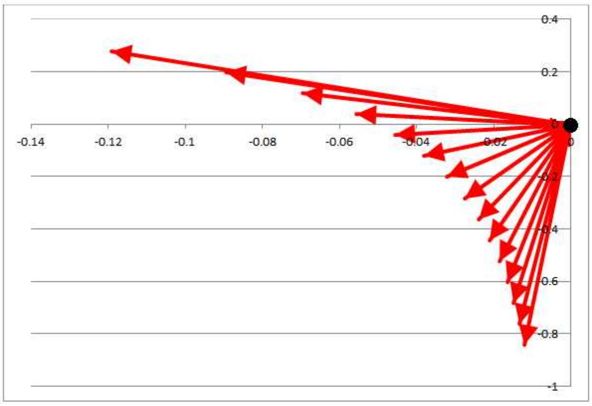
Suppose $\{A_k :\ k = 1, 2, 3...\}$ is a sequence of vectors in $\mathbf{R}^n$. First, our definition is equivalent to the following requirement:

▶ for each real number $\varepsilon > 0$, there exists a number N such that, for every natural number $k > N$, we have
$$||A_k - A|| < \varepsilon\,.$$

Definition 1.13.4: convergent sequence

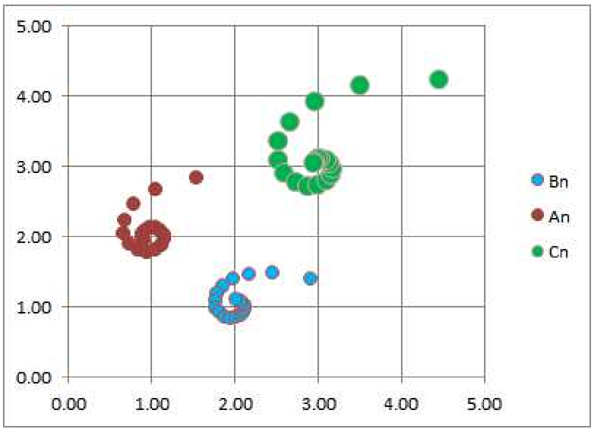
We say that a sequence of $\{A_k : \ k = 1, 2, 3...\}$ of vectors in $\mathbf{R}^n$ *converges* to another vector A in $\mathbf{R}^n$, called the *limit* of the sequence, if the following condition is satisfied:
$$||A_k - A|| \to 0 \ \text{ as } k \to \infty\,.$$

This limit is denoted as follows:
$$A_k \to A \ \text{ as } k \to \infty\,,$$
or
$$A = \lim_{k \to \infty} A_k\,.$$



We consider now the algebra of sequences and limits. Limits behave well with respect to the vector operations. Below we assume that the sequences are defined on the same set of integers.

We start with addition.



To graphically add two sequences, we plot parallelograms. Then, the diagonals of these parallelograms form the new sequence. Now, if either sequence converges to 0, then so do these diagonals.

Theorem 1.13.5: Sum Rule

$$\boxed{A_k \to 0 \ \text{ and } \ B_k \to 0 \implies A_k + B_k \to 0}$$

Proof.

From the assumption and the definition it follows:

$$||A_k|| \to 0 \quad \text{and} \quad ||B_k|| \to 0 \,.$$

Since these two *numerical* sequences converge to 0, then so does their sum:

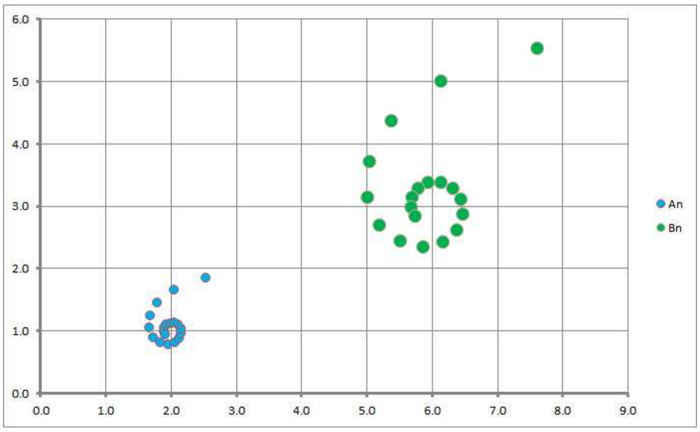
$$||A_k|| + ||B_k|| \to 0 \,,$$

as we know from the *Sum Rule for numerical sequences* in Volume 2 (Chapter 2DC-1). From the *Triangle Inequality* we conclude:

$$||A_k + B_k|| \le ||A_k|| + ||B_k|| \to 0 \,.$$

Therefore, by definition $A_k + B_k \to 0$.

Multiplying a sequence by a scalar simply stretches the whole picture uniformly.



Theorem 1.13.6: Constant Multiple Rule

$$\boxed{A_k \to 0 \implies cA_k \to 0 \text{ for any real } c}$$

Proof.

From the assumption and the definition it follows:

$$||A_k|| \to 0 \,.$$

Since this is a *numerical* sequence that converge to 0, then so does its multiple:

$$|c|\, ||A_k|| \to 0 \,,$$

as we know from the *Constant Multiple Rule for numerical sequences* in Volume 2 (Chapter 2DC-1). Then:

$$||cA_k|| = |c|\, ||A_k|| \to 0 \,.$$

Therefore, by definition $cA_k \to 0$.

For more complex situations we need to use the fact that convergent sequences are *bounded*.

Theorem 1.13.7: Boundedness

$$A_k \to A \implies ||A_k|| < Q \text{ for some real } Q \,.$$

Proof.

The idea is that the tail of the sequence will fit into some ball around the limit; meanwhile, there are only finitely many terms left... Choose $\varepsilon = 1$. Then by definition, there is such N that for all $k > N$ we have:

$$||A_k - A|| < 1\,.$$

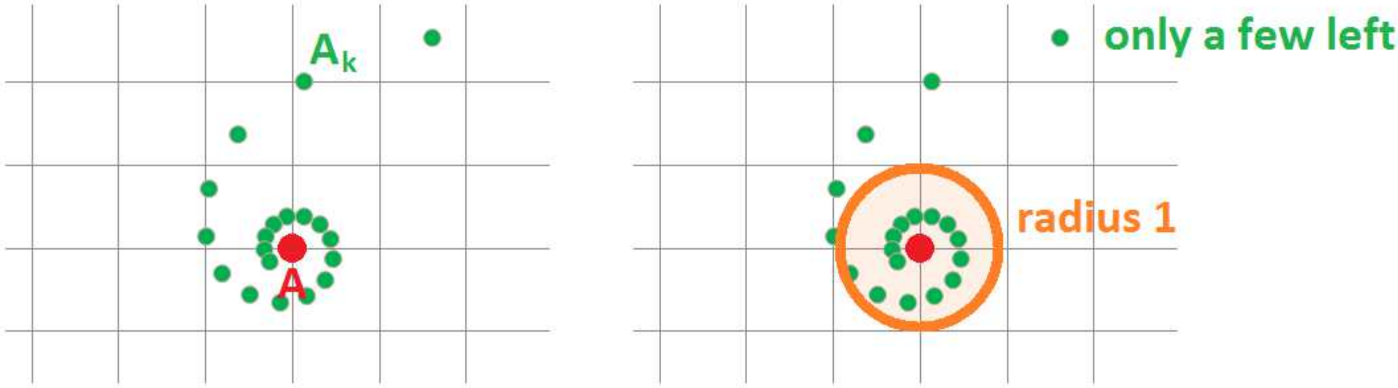
Then, we have:

$$\begin{aligned} ||A_k|| &= ||(A_k - A) + A|| && \text{By Triangle Inequality.}\\ &\le ||A_k - A|| + ||A|| && \text{By inequality above.}\\ &< 1 + ||A||. \end{aligned}$$

To finish the proof, we choose:

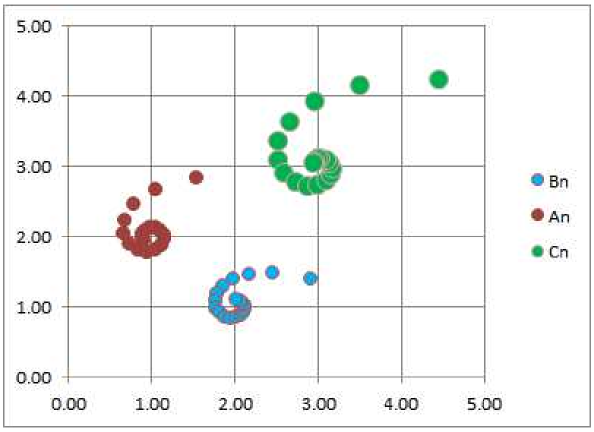
$$Q = \max\{||A_1||,\ \ldots,\ ||A_k||,\ 1 + ||A||\}\,.$$

The proof is illustrated below:



The converse isn't true: Not every bounded sequence is convergent. We will show later that, with an extra condition, bounded sequences do have to converge...

We are now ready for the general results on the algebra of limits.



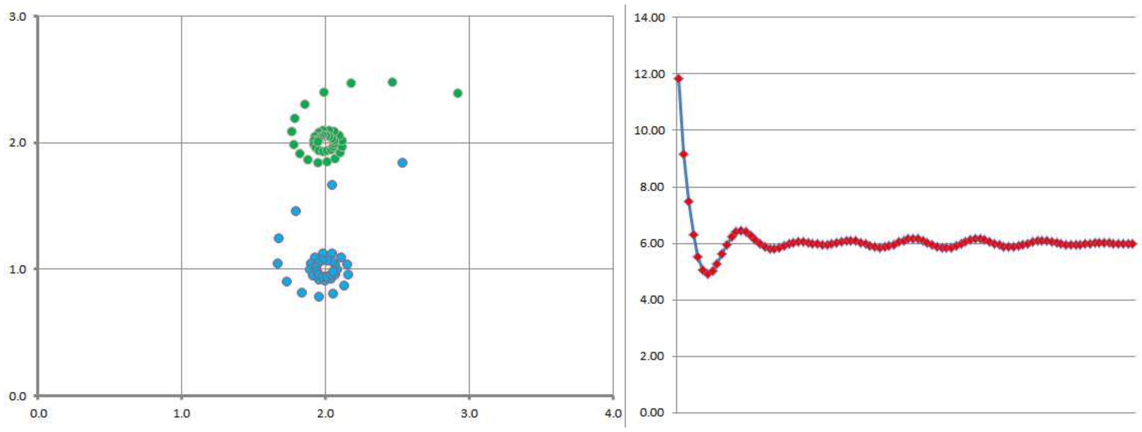
Theorem 1.13.8: Sum Rule

If sequences A_k, B_k converge then so does $A_k + B_k$, and

$$\boxed{\lim_{k\to\infty}(A_k + B_k) = \lim_{k\to\infty} A_k + \lim_{k\to\infty} B_k.}$$

Proof.

Suppose
$$A_k \to A \ \text{ and } \ B_k \to B\,.$$
Then,
$$||A_k - A|| \to 0 \ \text{ and } \ ||B_k - B|| \to 0\,.$$
We compute:
$$\begin{aligned} ||(A_k + B_k) - (A + B)|| &= ||(A_k - A) + (B_k - B)|| && \text{By Triangle Inequality.} \\ &\le ||A_k - A|| + ||B_k - B|| \\ &\to 0 + 0 && \text{By SR for numerical sequences.} \\ &= 0. \end{aligned}$$
Then, by the last theorem, we have:
$$||(A_k + B_k) - (A + B)|| \to 0\,.$$
Then, by the first theorem, we have:
$$A_k + B_k \to A + B\,.$$



Theorem 1.13.9: Dot Product Rule

If sequences A_k, B_k converge then so does $A_k \cdot B_k$, and we have:
$$\boxed{\lim_{k\to\infty} (A_k \cdot B_k) = \Big(\lim_{k\to\infty} A_k\Big) \cdot \Big(\lim_{k\to\infty} B_k\Big)}$$

Exercise 1.13.10

Prove the theorem.

The following is a corollary.

Theorem 1.13.11: Constant Multiple Rule

If sequence A_k converges then so does cA_k for any real c, and we have:
$$\boxed{\lim_{k\to\infty} c\,A_k = c \cdot \lim_{k\to\infty} A_k}$$

Warning!

It is considered a serious error if you use the conclusion (the formula) one of these rules without verifying the conditions (the convergence of the sequences involved).

We represent the *Sum Rule* as a diagram:

$$\begin{tikzcd}
A_k, B_k \arrow[r, "\lim"] \arrow[d, "+"] \arrow[dr, phantom, "SR"] & A, B \arrow[d, "+"] \\
A_k + B_k \arrow[r, "\lim"] & \lim(A_k + B_k) = A + B
\end{tikzcd}$$

In the diagram, we start with a pair of sequences at the top left and then we proceed in two ways:

- right: take the limit of either, then down: add the results; or
- down: add them, then right: take the limit of the result.

The result is the same!

1.14. *Partitions of the Euclidean space*

In Volumes 2 and 3, we used partitions of intervals, as well as the whole real line, in order to study *incremental change*. This time, we need partitions of the n-dimensional Euclidean space. The building blocks will come from partitions of the axes.

For dimension 2, these are *rectangles*. An interval in the x-axis:

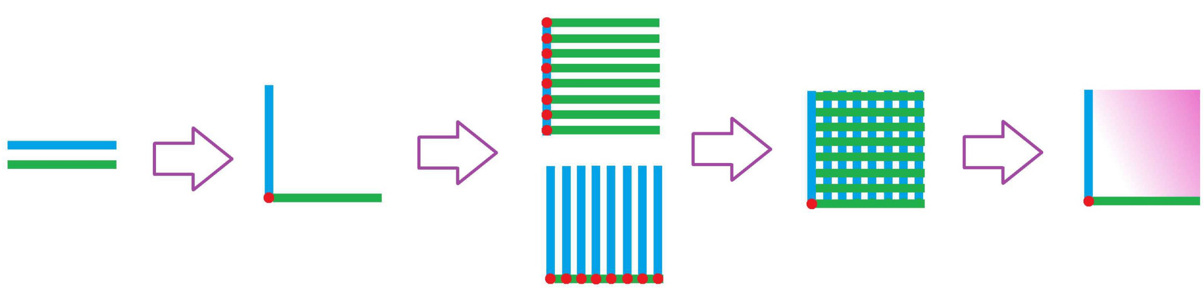
$$[a,b] = \{x :\ a \le x \le b\},$$

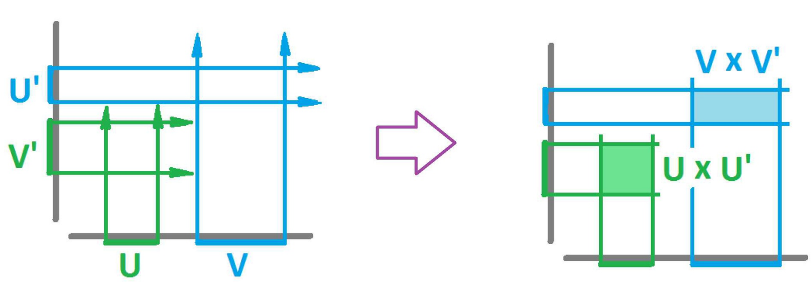
and an interval in the y-axis:

$$[c,d] = \{y :\ c \le y \le d\},$$

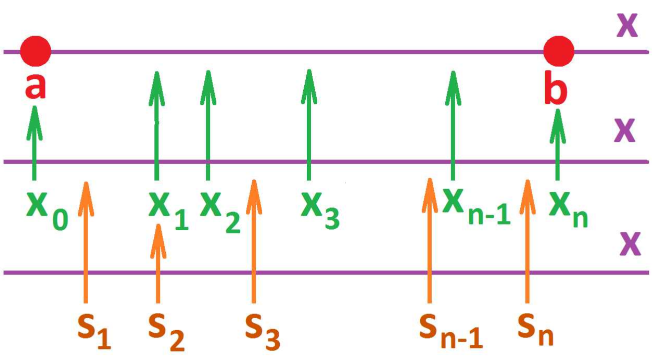
make a rectangle in the xy-plane:

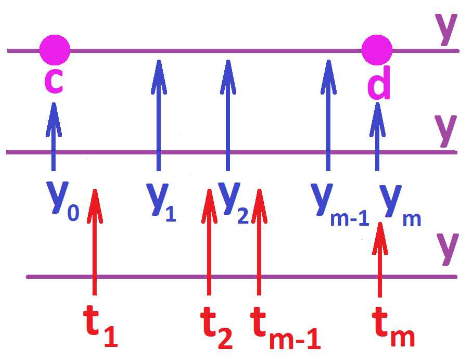
$$[a,b] \times [c,d] = \{(x,y) :\ a \le x \le b,\ c \le y \le d\}.$$





A *partition of the rectangle* $[a, b] \times [c, d]$ is made of smaller rectangles constructed in the same way as above. Suppose we have partitions of the intervals $[a, b]$ in the x-axis and $[c, d]$ in the y-axis:





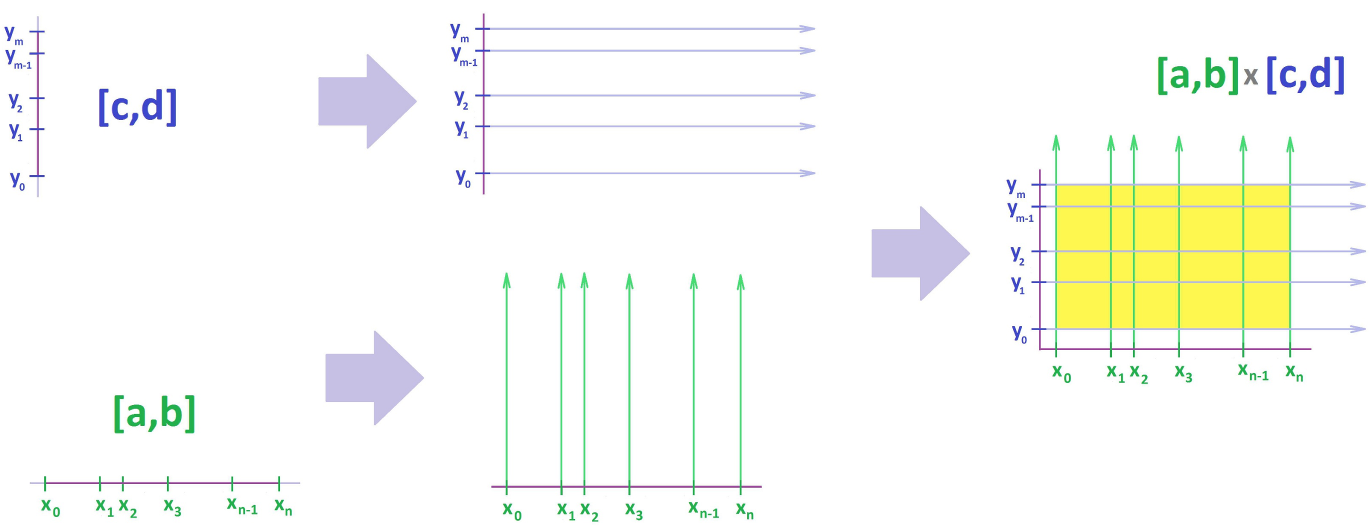
We start with a partition of an interval $[a, b]$ in the x-axis into n intervals:

$$[x_0, x_1],\ [x_1, x_2],\ \ldots,\ [x_{n-1}, x_n]\,,$$

with $x_0 = a,\ x_n = b$. Then we do the same for y. We partition an interval $[c, d]$ in the y-axis into m intervals:

$$[y_0, y_1],\ [y_1, y_2],\ \ldots,\ [y_{m-1}, y_m]\,,$$

with $y_0 = c,\ y_n = d$.

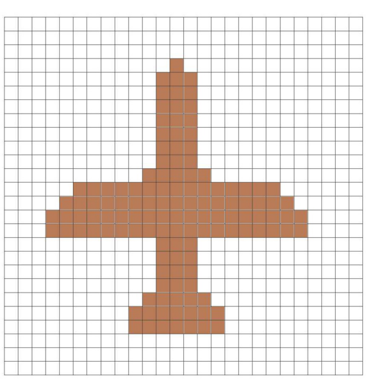


The lines $y = y_j$ and $x = x_i$ create a partition of the rectangle $[a, b] \times [c, d]$ into smaller rectangles $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$. The points of intersection of these lines,

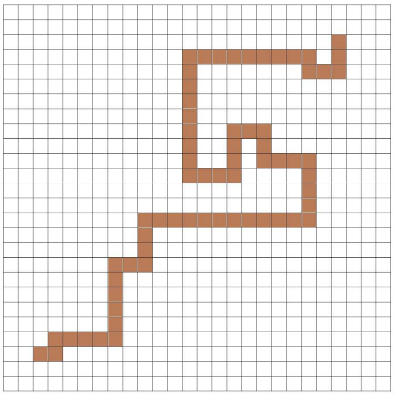
$$X_{ij} = (x_i, y_j),\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\,,$$

will be called the *nodes* of the partition. So, there are nodes and there are rectangles (tiles); is that it?

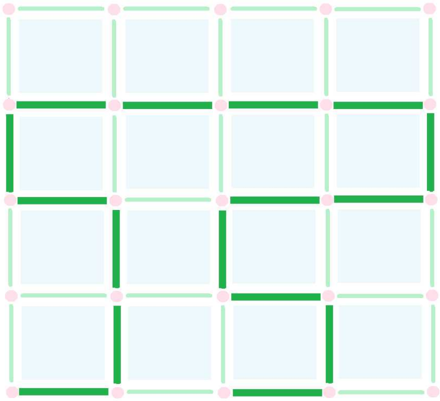
This is how an object can be represented with tiles, or pixels:



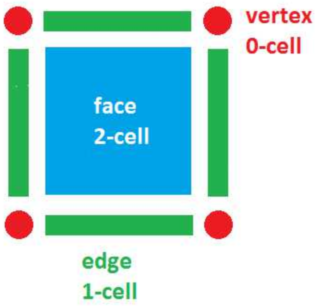
Now, are *curves* also made of tiles? Such a curve would look like this:



If we look closer, however, this "curve" isn't a curve in the usual sense; it's thick! The correct answer is: *curves are made of edges* of the grid:



We have discovered that we need to include, in addition to the squares, the "thinner" cells as additional building blocks. The complete decomposition of the pixel is shown below; the edges and vertices are shared with adjacent pixels:



Example 1.14.1: dimension 1

We start with dimension $n = 1$:

In this simplest of partitions, the cells are:

- A node, or a 0-cell, is $\{k\}$ with $k = ... - 2, -1, 0, 1, 2, 3, ...$
- An edge, or a 1-cell, is $[k, k+1]$ with $k = ... - 2, -1, 0, 1, 2, 3, ...$
- And, 1-cells are attached to each other along 0-cells.

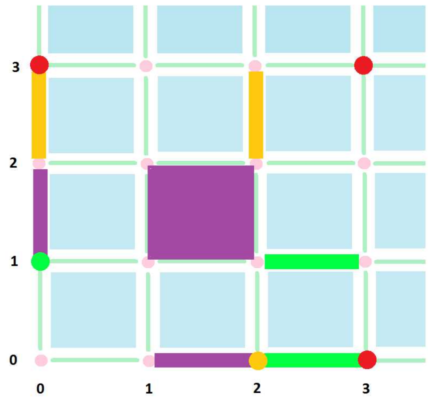
Example 1.14.2: dimension 2

For the dimension $n = 2$ grid, we define cells for all integers k, m as products:

- A node, or a 0-cell, is $\{k\} \times \{m\}$.
- An edge, or a 1-cell, is $\{k\} \times [m, m+1]$ or $[k, k+1] \times \{m\}$.
- A square, or a 2-cell, is $[k, k+1] \times [m, m+1]$.

We also have:

- The 2-cells are attached to each other along 1-cells.
- And, still, the 1-cells are attached to each other along 0-cells.



Cells shown above are:

- 0-cell $\{3\} \times \{3\}$,
- 1-cells $[2, 3] \times \{1\}$ and $\{2\} \times [2, 3]$,
- 2-cell $[1, 2] \times [1, 2]$.

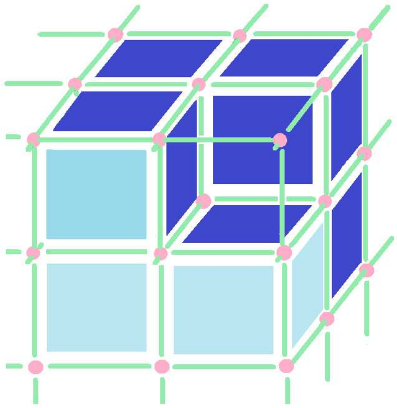
Similarly for dimension 3, we have *boxes*. Intervals in the x-, y-, and z-axes:

$$[a, b] = \{x : \ a \leq x \leq b\}, \ [c, d] = \{y : \ c \leq y \leq d\}, \ [p, q] = \{z : \ p \leq z \leq q\},$$

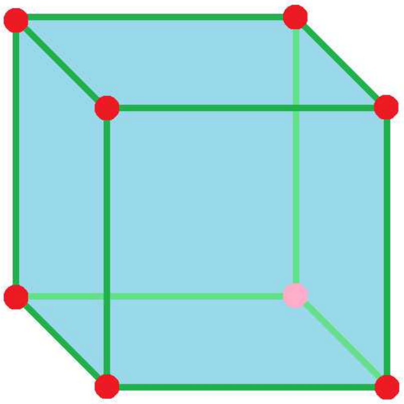
make a box in the xyz-space:

$$[a, b] \times [c, d] \times [p, q] = \{(x, y) : \ a \leq x \leq b, \ c \leq y \leq d, \ p \leq z \leq q\}.$$

In dimension 3, *surfaces are made of faces* of our boxes; i.e., these are tiles:



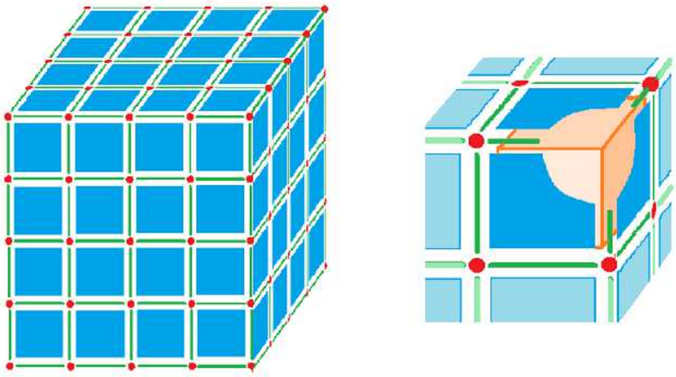
The cell decomposition of the box follows and here, once again, the faces, edges, and vertices are shared:



Example 1.14.3: dimension 3

For all integers i, m, k, we have:

- A node, or a 0-cell, is $\{i\} \times \{m\} \times \{k\}$.
- An edge, or a 1-cell, is $\{i\} \times [m, m+1] \times \{k\}$, etc.
- A square, or a 2-cell, is $[i, i+1] \times [m, m+1] \times \{k\}$, etc.
- A cube, or a 3-cell, is $[i, i+1] \times [m, m+1] \times [k, k+1]$.



Thus, our approach to decomposition of space, in any dimension, boils down to the following:

> ▶ *The n-dimensional space is composed of cells in such a way that k-cells are attached to each other along $(k-1)$-cells, $k = 1, 2, ..., n$.*

The examples show how the n-dimensional Euclidean space is decomposed into 0-, 1-, ..., n-cells in such a way that

- n-cells are attached to each other along $(n-1)$-cells.
- $(n-1)$-cells are attached to each other along $(n-2)$-cells.
- ...
- 1-cells are attached to each other along 0-cells.

What are those cells exactly?

Definition 1.14.4: cell

In the n-dimensional space, $\mathbf{R}^n$, a *cell* is the subset given by the product with n components:

$$P = I_1 \times ... \times I_n \,,$$

with its kth component is either

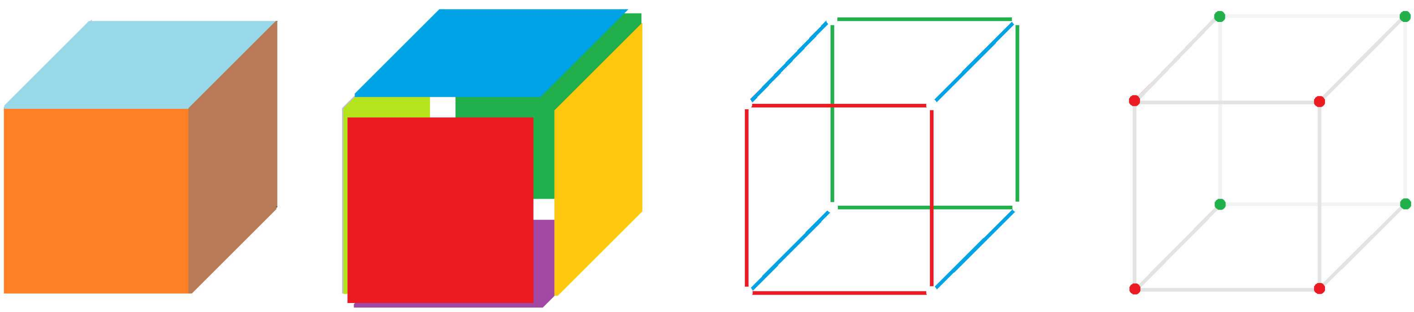
- *a closed interval* $I_k = [x_k, x_{k+1}]$, or
- *a point* $I_k = \{x_k\}$.

The cell's *dimension* is equal to m, and it is also called an m-cell, when there

are m edges and $m-n$ vertices on this list. Replacing one of the edges in the product with one of its end-points creates an $(n-1)$-cell called a *boundary cell* of P.

Definition 1.14.5: face

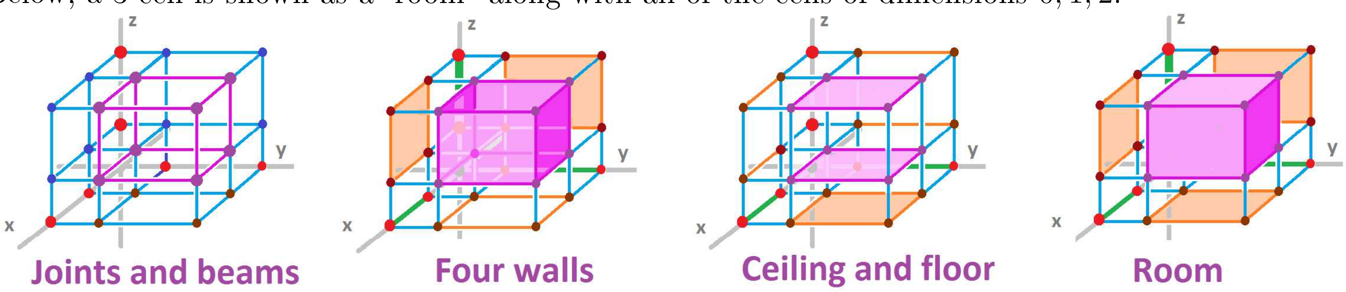
Replacing one of the edges in the product with one of its end-points creates an $(n-1)$-cell called a *face* of P. Replacing several edges with one of their end-points creates an k-cell, $k<n$, called a *boundary cell* of P.



Thus, partitions of the axes – into nodes and edges – create a partition of the whole space – into cells of all dimensions.

Example 1.14.6: 3-cell

Below, a 3-cell is shown as a "room" along with all of the cells of dimensions $0,1,2$:



They all come from the nodes and edges on the axes:

- 0: Each of the joints of the "beams" is the product of three nodes.
- 1: Each of the "beams" is the product of two nodes and an edge.
- 2: Each of the "walls", as well as the "floor" and the "ceiling", is the product of two edges and a node.
- 3: The "room" is the product of three edges.

The 2-cells here are the faces of the 3-cell, the 1-cells are the faces of the 1-cells, etc.

Definition 1.14.7: product of the partitions

Suppose each of the coordinate axes of the n-dimensional space, $\mathbf{R}^n$, has a partition. Then, the *product of the partitions* consists of the cells given by the product with n components:

$$P=I_1\times ...\times I_n\,,$$

with its kth component is either

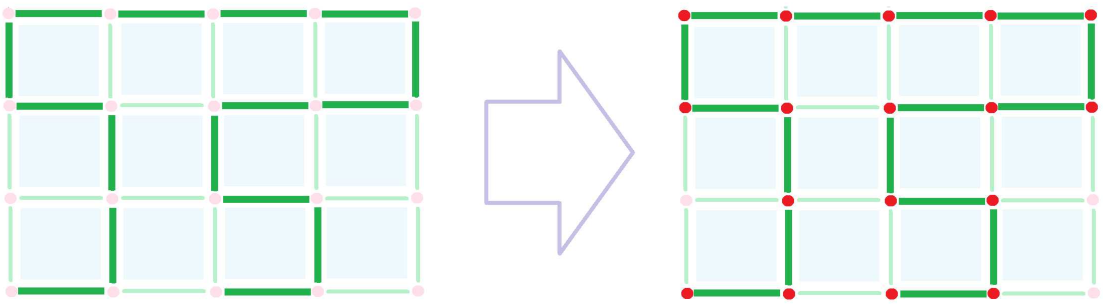
- *an edge* $I_k=[x_k,x_{k+1}]$, or
- *a node* $I_k=\{x_k\}$,

in the partition of the kth axis. This combination of cells is called a *partition* in $\mathbf{R}^n$.

Definition 1.14.8: partition

Suppose we have a partition in $\mathbf{R}^n$ and suppose a subset D of $\mathbf{R}^n$ is the union of some of the cells in the partition. We say that this is a *partition* of D provided: if a cell is present, then so do all of its boundary cells.

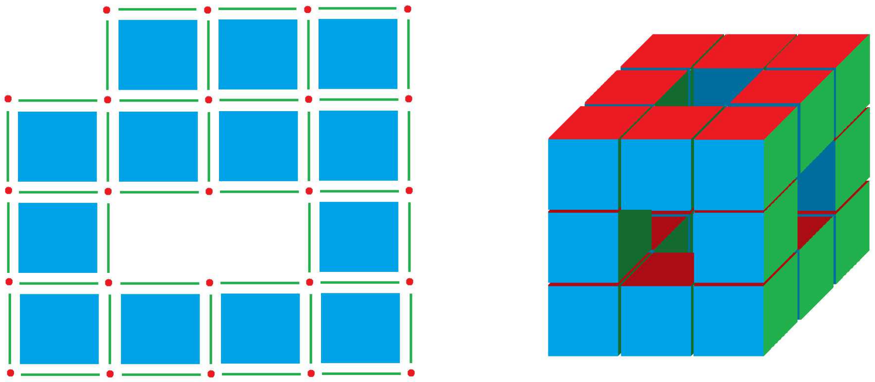
For example, *any* sequence of edges Q_i, $i = 1, ..., n$, of a partition can be seen as a curve. However, a partition of the curve also includes all of the end-points of the edges.



Then a *continuous* curve consists of a sequence of *consecutive* edges or, which is the same, of a sequence of *adjacent* nodes:

$$Q_i = P_{i-1}P_i \, .$$

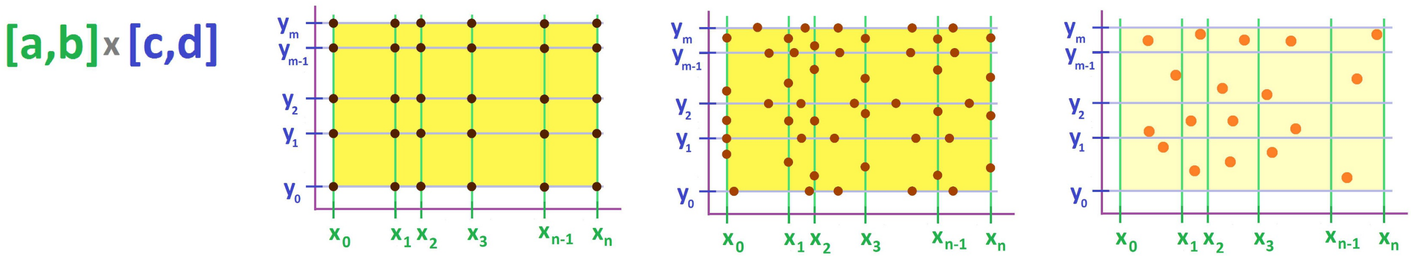
Furthermore, for a partition of a surface made of faces, we must also have all the edges and the nodes of these faces, and so on.



We will carry out all calculus constructions within these partitions.

1.15. Discrete forms

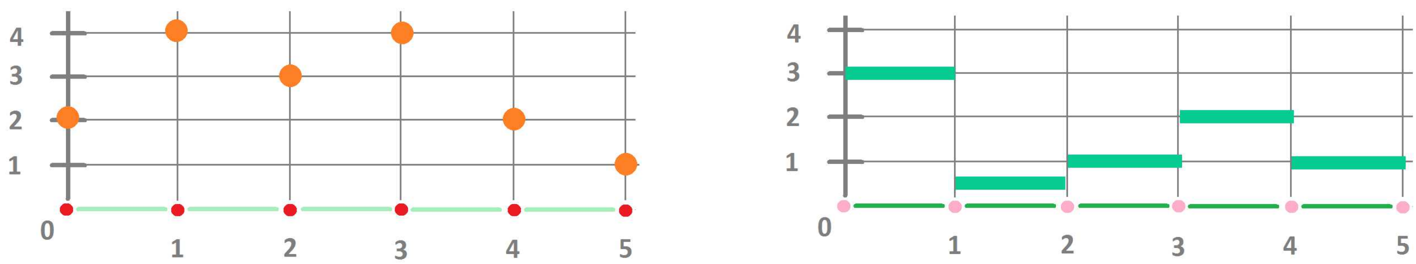
In Volumes 2 and 3, we assigned numbers to *points* within cells in the 1-dimensional case to represent such things as location – nodes or 0-cells – and velocity – secondary nodes or 1-cells. We will continue to do so. In fact, we will study functions defined at points located at the cells of a particular dimension m in a partition. Below we see $m = 0, 1, 2$:



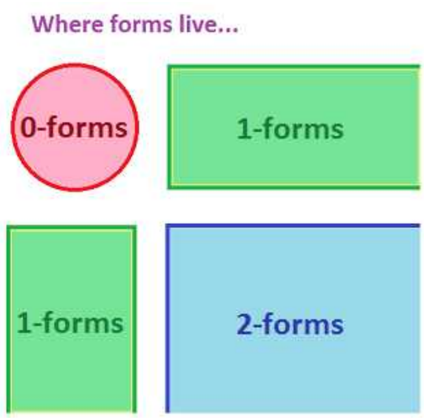
Firstly, these points – secondary, tertiary, etc. nodes – may be specified as a result of *sampling* a function defined on the whole region. Note that, in that case, one node may be shared by two adjacent cells.

Secondly, these points are used for mere *bookkeeping*. We then can choose them to be the end-points or corners or mid-points etc. In truth though, the quantities are assigned to the cells *themselves*. In other words, each cell is an input of these functions, as explained below.

Recall how we defined discrete forms for dimension 1: within each of the pieces of a partition of the line this function is unchanged; i.e., it's a single number. This is how we plot the graphs of 0- and 1-forms over $\mathbf{R}^1$:



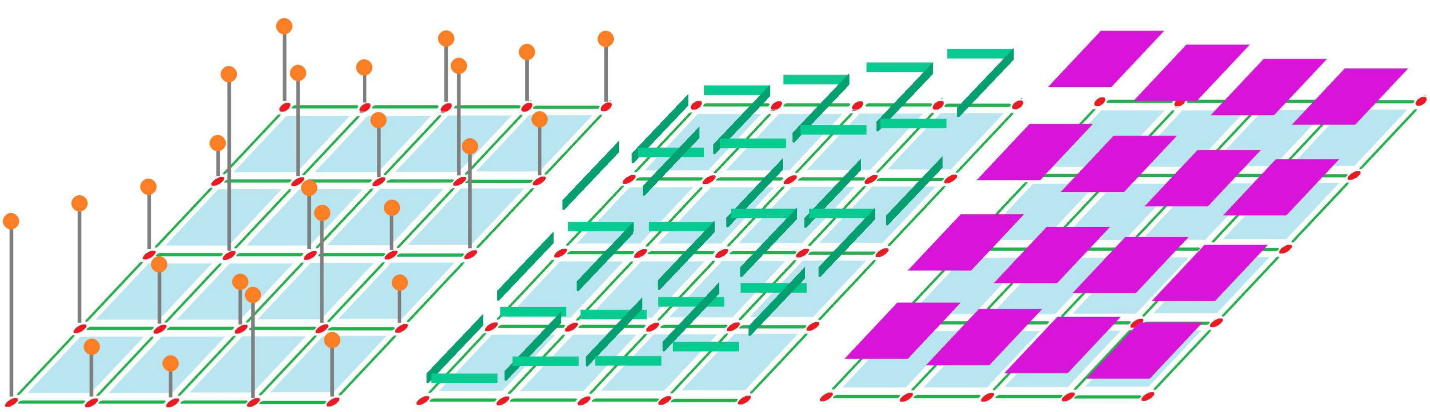
There are more types of cells in the higher dimensional spaces, but the idea remains:



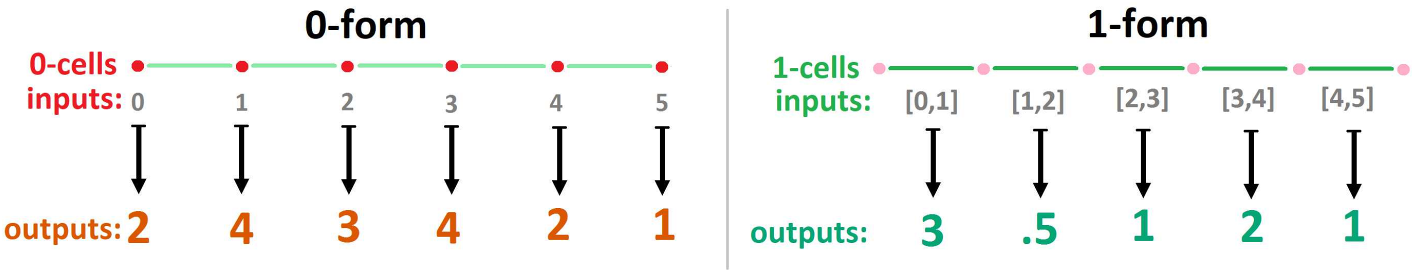
Definition 1.15.1: discrete form

A *discrete form of degree* k *over* $\mathbf{R}^n$, or simply a k-form, is a real-valued function defined on k-cells of $\mathbf{R}^n$.

And these are 0-, 1-, and 2-forms over $\mathbf{R}^2$:



To emphasize the nature of a form as a function, we can use arrows ($\mathbf{R}^1$):



Here we have two forms:

- a 0-form with $0 \mapsto 2,\ 1 \mapsto 4,\ 2 \mapsto 3, ...$; and
- a 1-form with $[0,1] \mapsto 3,\ [1,2] \mapsto .5,\ [2,3] \mapsto 1, ...$.

A more compact way to visualize is this:

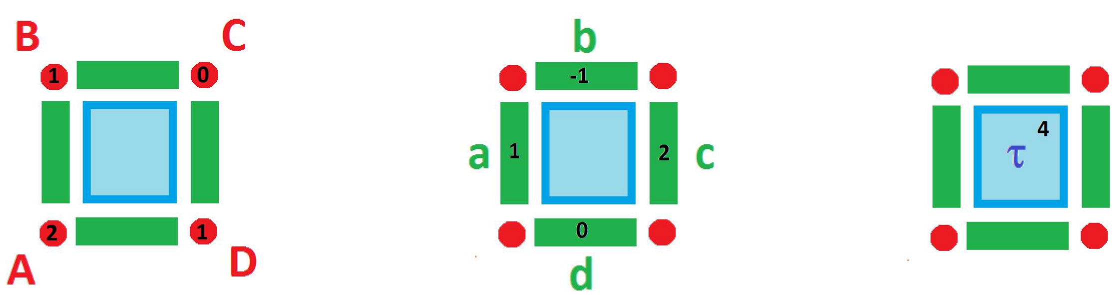
Here we have two forms:

- a 0-form q with $q(0) = 2,\ q(1) = 4,\ q(2) = 3, ...$; and
- a 1-form s with $s\Big([0,1]\Big) = 3,\ s\Big([1,2]\Big) = .5,\ s\Big([2,3]\Big) = 1, ...$.

We can also use letters to label the cells, just as before. Each cell is then assigned *two* symbols:

- one is its name (a latter) and
- The other is the value of the form at that location (a number).

This idea is illustrated for forms over $\mathbf{R}^1$ and $\mathbf{R}^2$ respectively:



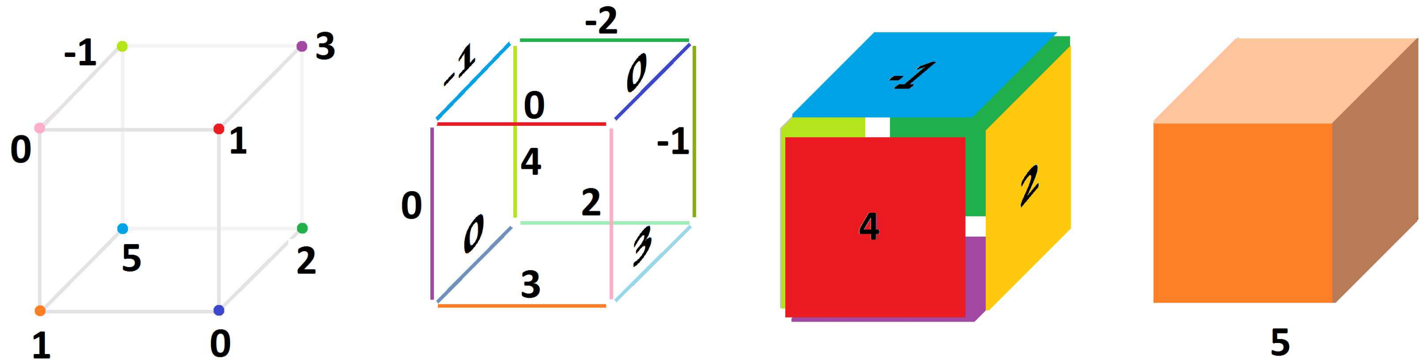
We have a 0-form q and a 1-form s in the former example:

- $q(A) = 2,\ q(B) = 4,\ q(C) = 3, ...$
- $s(AB) = 3,\ s(BC) = .5,\ s(CD) = 1, ...$

We also have a 2-form ϕ in the latter example:

- $q(A) = 2,\ q(B) = 1,\ q(C) = 0,\ q(D) = 1$
- $s(a) = 1,\ s(b) = -1,\ s(c) = 2, s(d) = 0$
- $\phi(\tau) = 4$

We can simply label the cells with numbers, as follows (in $\mathbf{R}^3$):



These forms may represent the following characteristics of a flow of a liquid:

- A 0-form: the pressure of the liquid at the joints of a system of pipes.
- A 1-form: the flow rate of the liquid along the pipe.
- A 2-form: the flow rate of the liquid across the membrane.
- A 3-form: the density of the liquid inside the box.

These forms will be used to study functions of several variables in the following chapters. However, the example of parametric curves – and especially motion in space – suggests that we may need the domain of these functions to be multi-dimensional. We saw a function defined, just like a 0-form, at the nodes of the partition of the line, but with values in $\mathbf{R}^2$, unlike a 0-form.

This is a generalization of the last definition.

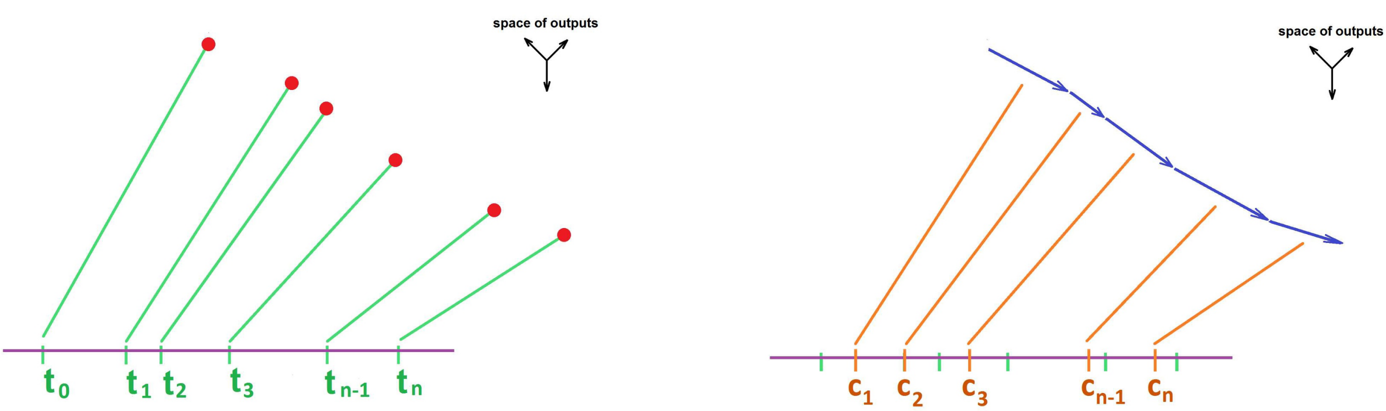
Definition 1.15.2: vector-valued discrete form

Suppose n and m are given. Then a *vector-valued discrete form F of degree k*, or simply a k-form, is a function defined on k-cells of $\mathbf{R}^n$ with values in $\mathbf{R}^m$.

As you can see, we will be using capital letters for vector-valued forms in accordance with our convention.

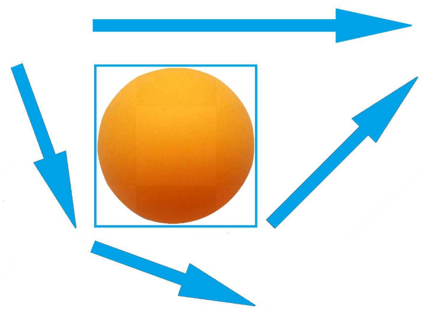
Note that discrete forms do *not* exactly match our list of functions: numerical functions, parametric curves, vector fields, and functions of several variables. From the same domain, we pick cells of different dimensions producing forms of different degrees.

This is an illustration of two vector-valued forms: a 0-form and a 1-form (for the latter, the vectors have to be moved to put the starting points at the origin); i.e., $n=1$ and $m=2$:



The former may represent the locations and the latter the velocities. Both can be seen as parametric curves.

Next is an illustration of a real-valued and a vector-valued 1-forms; i.e., $n=2$, $m=1$ and $n=2$, $m=2$ respectively:



The former may represent a flow of water along a system of pipes and the latter the same flow with possible leaks. Both can be seen as vector fields.

The algebra of vectors allows us to reproduce the definitions from Volume 3 (Chapter 3IC-4) in the new, multi-dimensional in both input and output, context. We just assume that a space of inputs $\mathbf{R}^n$ with a partition and a space of outputs $\mathbf{R}^m$ are given.

We need one more generalization; we introduce *orientation of cells.*

About the 0-cells, we will simply allow them to appear with both positive and negative signs.

Definition 1.15.3: oriented 0-cell

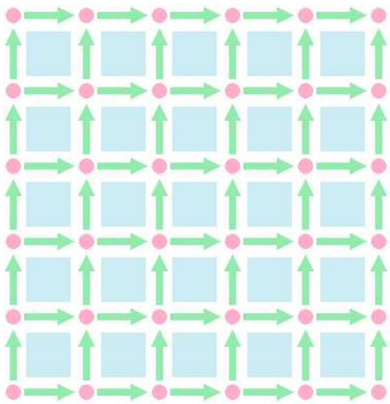
An *oriented* 0-*cell* (or node) A is a 0-cell of the partition with its sign specified: A or $-A$.

The choice will depend on the circumstances. The need for this will become clear shortly.

We have 1-cells defined as products of a single edge and several nodes. The order of nodes doesn't matter:

$$E = AB = BA,\ A = (a_1, ..., a_n),\ B = (b_1, ..., b_n) \implies a_i = b_i,\ i = 1, 2, ..., n, \quad \text{and} \quad a_k \neq b_k \text{ for some } k\,.$$

We would like to distinguish the two ways we can follow an edge.



Definition 1.15.4: oriented 1-cell

An *oriented* 1-*cell* (or edge) E is a 1-cell of the partition with the order of its two nodes specified: AB or BA. The cell is called *positively oriented* of the order of the nodes does with the directions of the axes:

$$E = AB,\ A = (a_1, ..., a_n),\ B = (b_1, ..., b_n) \implies a_k < b_k \text{ for some } k\,.$$

The cell is called *negatively oriented* of the order of the nodes goes against the directions of the axes:

$$E = AB,\ A = (a_1, ..., a_n),\ B = (b_1, ..., b_n) \implies a_k > b_k \text{ for some } k\,.$$

Negative edge

$$BA = -AB\,.$$

The orientation of higher-dimensional cells is addressed later.

The following is crucial:

Definition 1.15.5: boundary of a 0-cell

The *boundary of a 0-cell* is 0, denoted as follows:

$$\boxed{\partial A = 0}$$

The *boundary of a 1-cell* is the difference of its end-points denoted as follows:

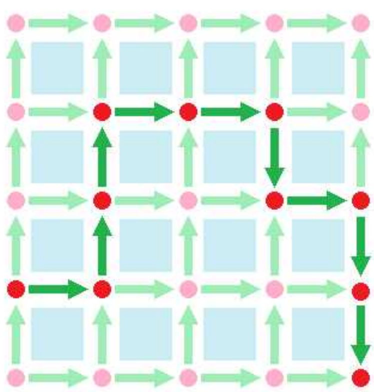
$$\boxed{\partial AB = B - A}$$

Warning!

While ∂A is a number, ∂AB is an algebraic combination of 1-cells.

Even though *any* sequence of edges $E_i,\ i = 1, ..., n$, is seen as a curve, a *continuous* curve consists of a sequence of *consecutive* oriented edges or, which is the same, of a sequence of *adjacent* nodes:

$$E_i = P_{i-1}P_i\,.$$

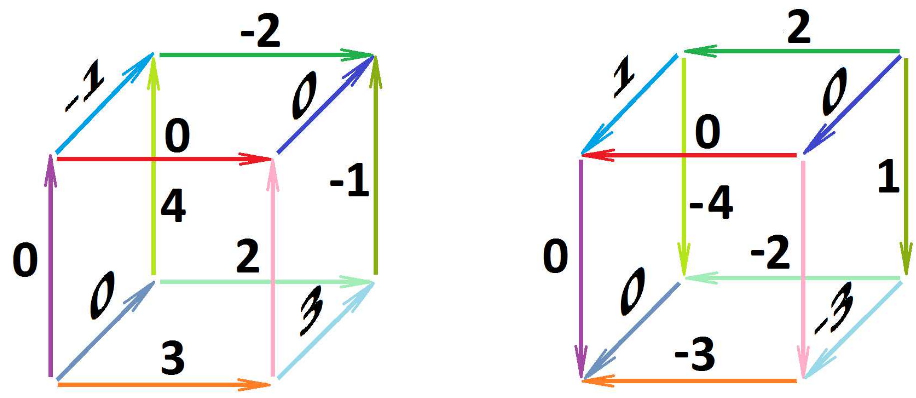


What is the sum of the boundaries of the edges of C?

Definition 1.15.6: boundary of a curve

The *boundary of a curve* C from A to B is $B - A$.

Recall that 0-forms are defined on the positively oriented 0-cells... and now they are instantly extended to the negatively oriented cells. Similarly, 1-forms are defined on both positively and negatively oriented 1-cells.



We amend our definition.

Definition 1.15.7: discrete form

Suppose n and m are given. Then a *real-valued or vector-valued discrete form* F *of degree* k, or simply a k-form, $k = 0, 1$, is a function defined on the positively and negatively oriented k-cells of $\mathbf{R}^n$ with values in $\mathbf{R}$ or $\mathbf{R}^m$ respectively so that:

$$F(-a) = -F(a)\,.$$

Now calculus.

Definition 1.15.8: difference

The *difference of a discrete* 0*-form* F is a discrete 1-form given by its values on each edge $E = AB$ of the partition:

$$\boxed{\Delta F\,(E) = F(B) - F(A)}$$

A 1-form is called *exact* when it is the difference of some 0-form.

The picture above may serve as an illustration of this concept.

Definition 1.15.9: sum

The *sum of a discrete* 0*-form* F along a collection Q of oriented nodes $N_1, N_2, ..., N_k$ is defined and denoted to be the following:

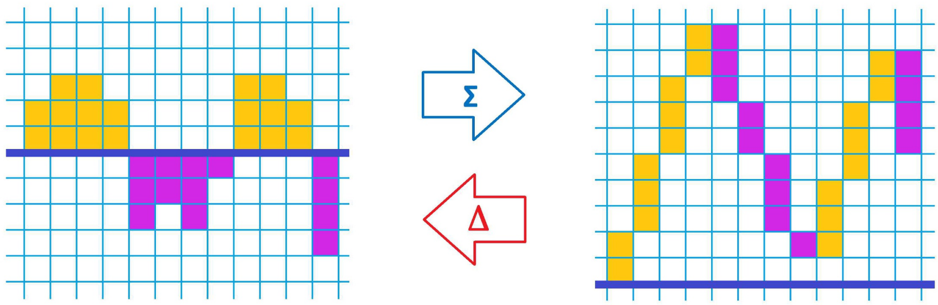
$$\boxed{\sum_Q F = F(N_1) + F(N_2) + ... + F(N_k)}$$

Definition 1.15.10: sum

The *sum of a discrete* 1*-form* G along a collection C of oriented edges $E_1, E_2, ..., E_k$ is defined and denoted to be the following:

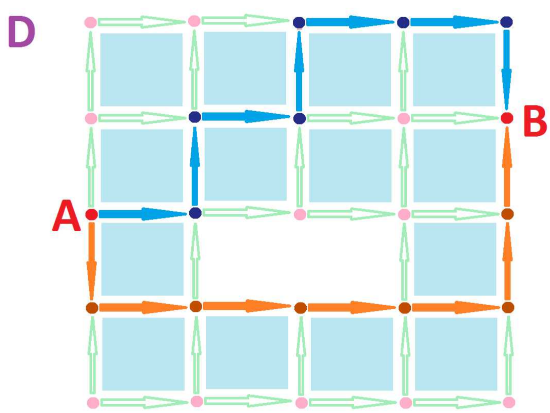
$$\boxed{\sum_C G = G(E_1) + G(E_2) + ... + G(E_k)}$$

The relation between the two operations remains the same as in the 1-dimensional case: they cancel each other.



Recall first that *any* sequence of edges $Q_i,\ i = 1, ..., n$, of a partition is seen as a curve, while a *continuous* curve consists of a sequence of *consecutive* edges or, which is the same, of a sequence of *adjacent* nodes:

$$Q_i = P_{i-1}P_i \,.$$



Theorem 1.15.11: Fundamental Theorem of Discrete Calculus of Degree 1

Suppose a partition of an n-cell in $\mathbf{R}^n$ is given. Suppose F is a discrete 0-form on this partition and suppose A is a node of the partition. Then, for each node X and any continuous curve from A to X, we have:

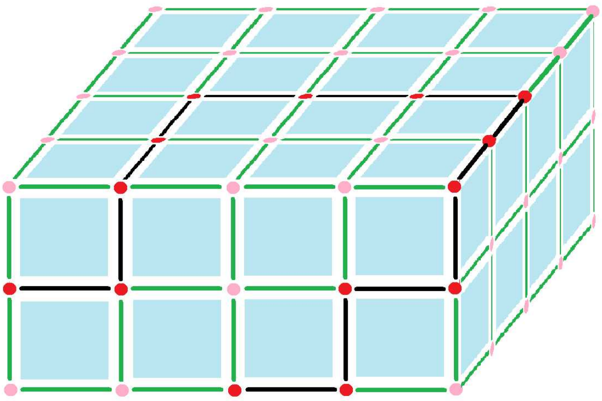
$$\boxed{\sum_C (\Delta F) = F(X) - F(A)}$$

Proof.

We just add all of these and cancel the repeated nodes:

$$\begin{array}{llllll}
\sum\limits_C G & = G(E_1) & +G(E_2) & +... & +G(E_k) \\
& = G(P_0P_1) & +G(P_1P_2) & +... & +G(P_{k-1}P_k) \\
& = \big[F(P_1) - F(P_0)\big] & +\big[F(P_2) - F(P_1)\big] & +... & +\big[F(P_k) - F(P_{k-1})\big] \\
& = -f(P_0) & & & +F(P_k) \\
& = F(X) - F(A)\,. & & &
\end{array}$$

In other words, if F is the difference over every edge, it is the difference over any continuous curve.



Corollary 1.15.12: Fundamental Theorem of Discrete Calculus of Degree 1

Under the conditions of the theorem, we have:

$$\boxed{\sum_{C}(\Delta F) = \sum_{\partial C} F}$$

But do they cancel each other in either order? Does this formula from Volume 1 (Chapter 1PC-1) still make sense:

$$\Delta\left(\sum_{C} G\right) = G(X)\,?$$

We will address this question in the following chapters.

Chapter 2: Parametric curves

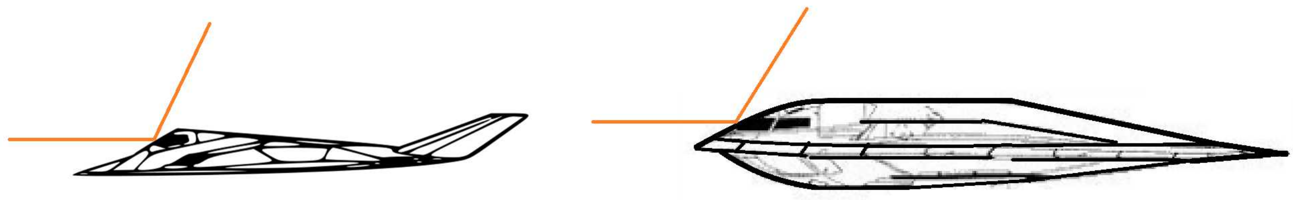
Contents

2.1. Parametric curves

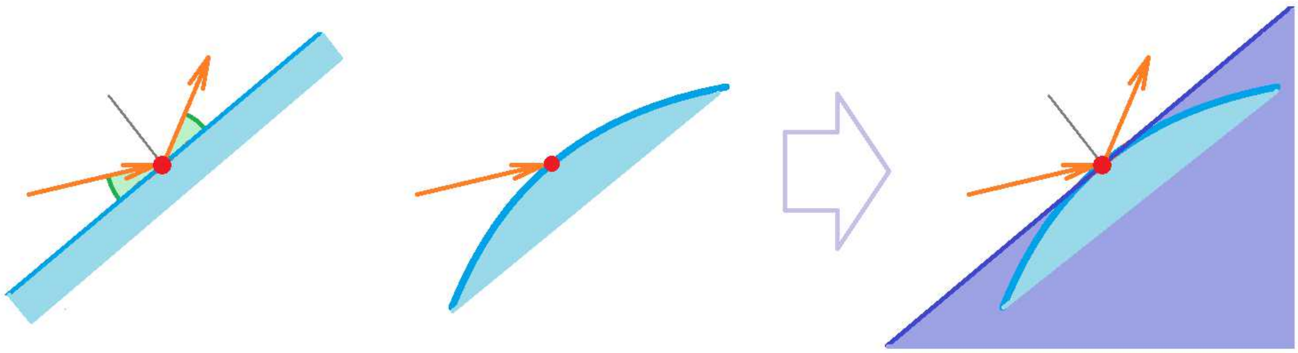
Recall that the *Tangent Problem* asks for a tangent line to a curve at a given point. It has been solved for the graphs of numerical functions in Volume 2 (Chapter 2DC-3). However, most of the curves in real life can't be represented as graphs. We have to look at parametric curves. The examples are familiar.

Example 2.1.1: radar

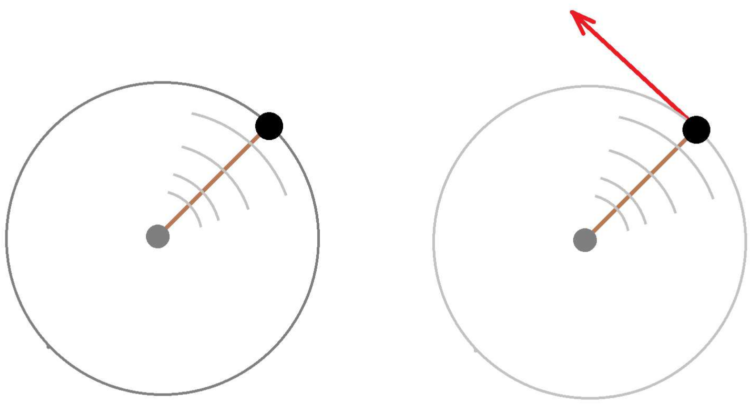
In which direction a radar signal will bounce off a plane when the surface of the plane is curved?



In what direction will light bounce off a curved mirror?



In what direction would a rock released from a sling go?



And so on.

Example 2.1.2: elliptical room

Let's confirm that the sound originating from one focus of an elliptical room will bounce off the wall to pass through the other focus.
This way one can listen to a conversation at the other focus even if there are obstacles between them.

Now, we study functions. They all have a single variable input and a single variable output. What varies is the *nature* of these two variables.

A parametric curve is such a function:

$$F : t \mapsto F(t)\,.$$

- The single independent variable is a *real number*, t.
- The single dependent variable is *multi-dimensional*, a point or a vector in $\mathbf{R}^n$.

For example for $n = 2$, we may have:

$$F : t \mapsto X = (f(t),\ g(t))\,,$$

or

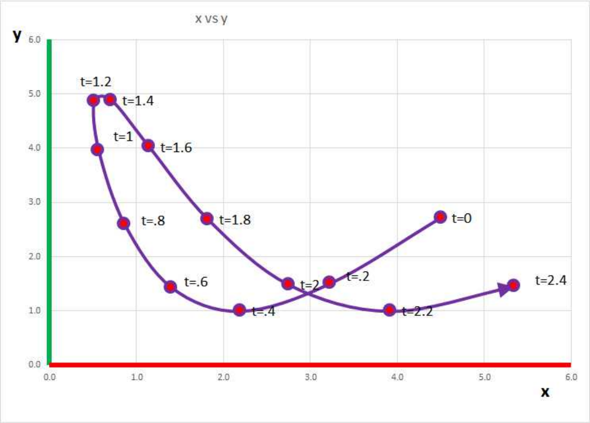
$$F : t \mapsto OX =< f(t),\ g(t) > \ .$$

As we know, the former point, X, is the end of the latter vector, OX. In either case, this is just a combination of two function of the same independent variable.

The go-to metaphor for parametric curves is *motion*:

- t is thought of as time,
- $(f(t), g(t))$ is thought of as the position in space at time t.

We plot each (x, y) and label each with the corresponding values of t:



These values of t may come as the nodes of a partition of an interval $[a, b]$ or run through the whole interval.

Example 2.1.3: straight line

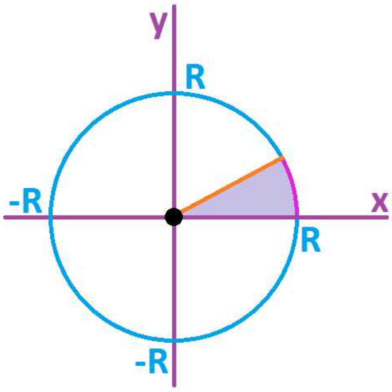
Linear motion is represented as follows:

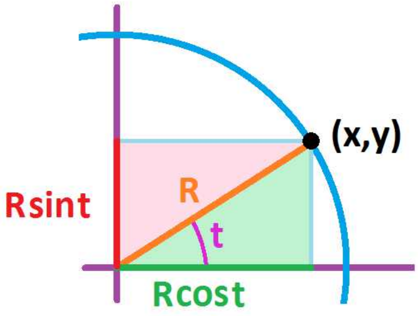
- $F(t) = V \cdot t$, a motion along a vector V (constant speed) from the origin
- $F(t) = V \cdot t + B$, same but with B the starting location
- $F(t) = -V \cdot t = V \cdot (-t)$, backward
- $F(t) = V \cdot (t^2)$, accelerated

Example 2.1.4: circle

The distance to some point remains the same, say, 1, and the implicit equation is

$$x^2 + y^2 = 1\,.$$



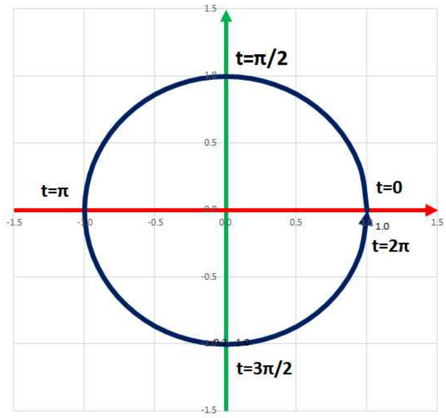


The "simple" circular motion happens when the *angular velocity* is constant: the object turns the same angle per unit of time:

$$F(t) = < \cos t,\ \sin t > \ .$$

We substitute $x = \cos t$ and $y = \sin t$ into the equation and use the *Pythagorean Theorem* to prove that this is indeed the unit circle:

$$x^2 + y^2 = (\cos t)^2 + (\sin t)^2 = 1\,.$$



There are other ways to travel around a circle of course.

- Circling 0 but moving backward:

$$F(t) =< \cos(-t),\ \sin(-t) > \ .$$

- Circling 0 but along radius r:

$$F(t) = r\cdot < \cos t,\ \sin t > \ .$$

- Circling point (p, q):

$$F(t) = (p, q) + r\cdot < \cos t,\ \sin t > \ .$$

- For accelerated motion, just change how fast the time goes:

$$F(t) =< \cos t^2,\ \sin t^2 > \ .$$

Example 2.1.5: ellipse

The implicit equation of the ellipse is

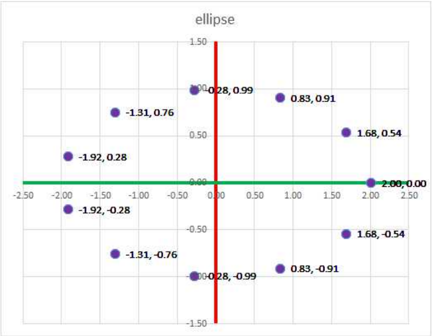
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\,,$$

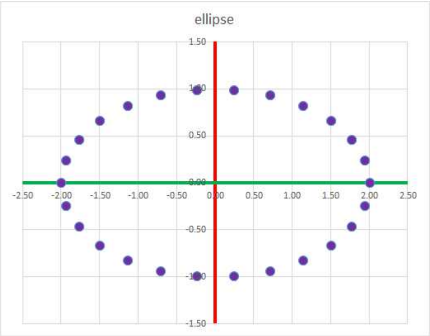
for some non-zero constant numbers a, b. Now, the "simple" elliptic motion is similar to circular motion happens when the angular velocity is constant but the distance varies.

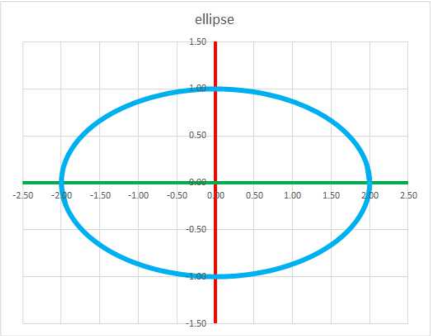
$$F(t) =< a\cos t,\ b\sin t > \ .$$

We substitute $x = a\cos t$ and $y = b\sin t$ into the equation and use the *Pythagorean Theorem* to prove that this is indeed the ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{(a\cos t)^2}{a^2} + \frac{(b\sin t)^2}{b^2} = \cos^2 t + \sin^2 t = 1\,.$$







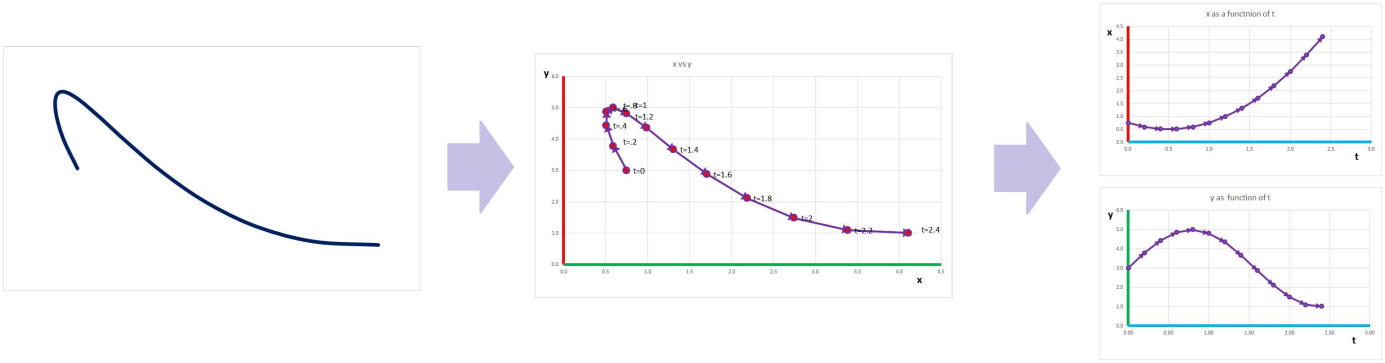
Other ways to travel around an ellipse are shown below. Circling 0 but moving backward:

$$F(t) =< a\cos(-t),\ b\sin(-t) > \ ,$$

circling point (p, q):

$$F(t) = (p, q) + r\cdot < a\cos t,\ b\sin t > \ .$$

Curves – given as graphs or implicitly – can be *parametrized*. It is as if we need to get the shape of the road and achieve that by driving along the road while recording the time and the location.



Definition 2.1.6: parametrization of curve

A parametric curve $X = X(t)$ is called a *parametrization of a curve* C in $\mathbf{R}^n$ when the path of X coincide with the curve C.

In the above examples we showed a variety of way to parametrize lines, circles, and ellipses.

Theorem 2.1.7: Reparametrization

If the function t=g(s) is one-to-one and onto, then the two parametric curves

$$X = F(t) \quad \text{and} \quad X = F(g(s))$$

are parametrizations of the same curve.

Definition 2.1.8: standard parametrization of ellipse

The parametrization of the ellipse that uses the angle of the line from the origin as the parameter is called the *standard parametrization of the ellipse*.

The graphs of functions are especially easy to parametrize.

Theorem 2.1.9: Parametrization of Graph

The parametric curve $x = t,\ y = f(t)$ *is a parametrization of the graph of* $y = f(x)$.

It is as if we are moving east at a constant speed...

Definition 2.1.10: standard parametrization of graph

For $n = 2$, we call the parametric curve

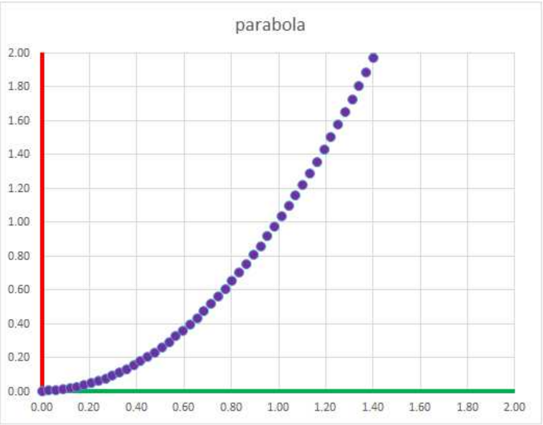
$$\begin{cases} x = t, \\ y = f(t), \end{cases}$$

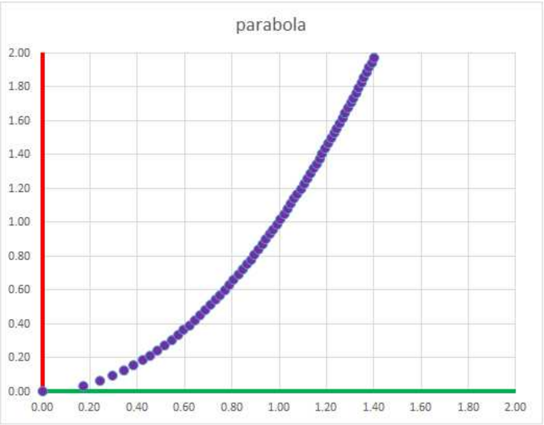
the *standard parametrization of the graph* $y = f(x)$.

Example 2.1.11: parabola

The standard parametrization of the parabola $y = x^2$ is (left):

$$F(t) = (t, t^2)\,.$$



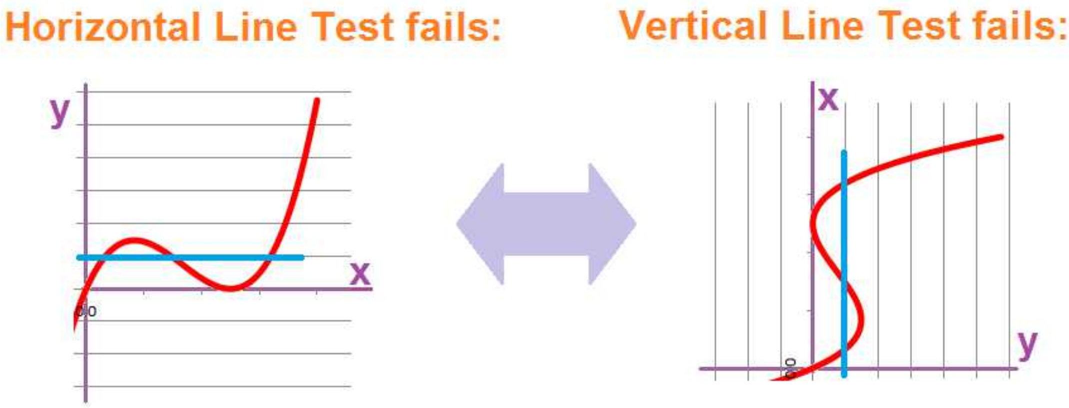


Meanwhile this parametrization of the parabola $x = \sqrt{y}$ is different (right):

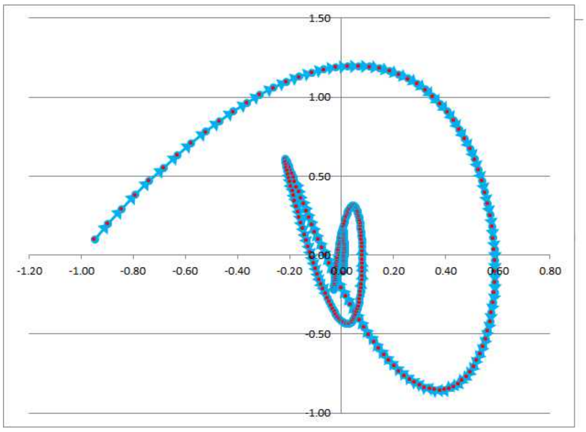
$$G(t) = (\sqrt{t}, t)\,.$$

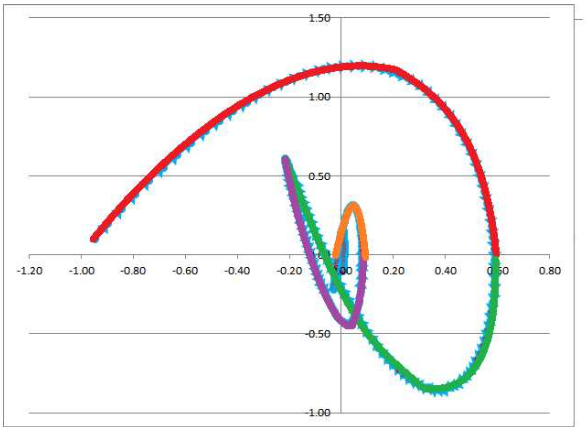
Conversely, we can sometimes represent the path of a plane parametric curve as the graph of a function. This function is

- $y = f(x)$ if the path satisfies the *Horizontal Line Test*, or
- $x = g(y)$ if the path satisfies the *Vertical Line Test*.



The circle fails both! However, it can be "de-parametrized" piece-by-piece (top-bottom or left-right halves). This would be a challenging task for more complex curves:





Representing a plane curve as a parametric curve from the beginning is a strongly preferred approach.

Algebraically, we try to *eliminate the parameter* by solving for t followed by substitution:

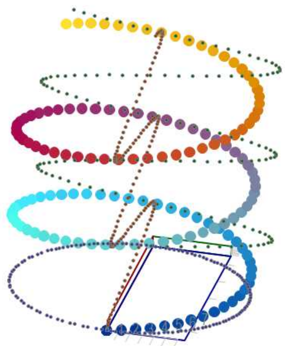
$$\begin{cases} x = t^3 \\ y = \sin t \end{cases} \implies t = \sqrt[3]{x} \implies y = \sin\left(\sqrt[3]{x}\right)\,.$$

The scheme works only when one of the component functions is *one-to-one*.

The motion may also be in the 3-dimensional space:

- t is thought of as time,
- $(f(t), g(t), h(t))$ is the position in space at time t.

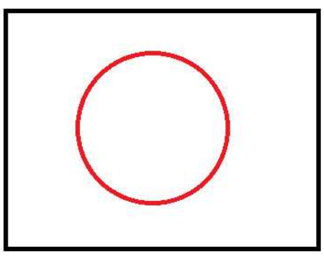
We plot each (x, y, z) and label each with the corresponding values of t.

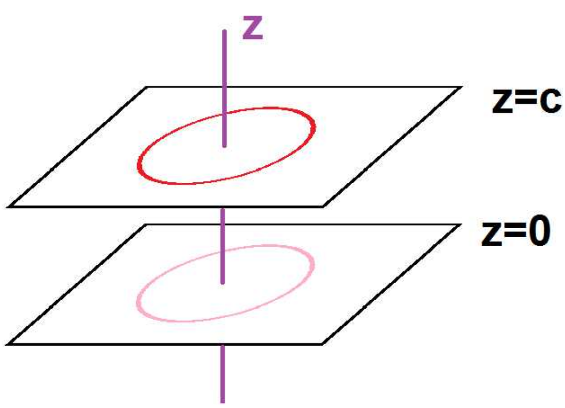


t	labels	x-value	y-value	z-value
0.00	1	1.00	0.00	0.00
0.00	2	0.99	0.12	0.01
0.00	3	0.97	0.24	0.02
0.00	4	0.94	0.35	0.04
0.00	5	0.89	0.46	0.05
0.01	6	0.83	0.56	0.06
0.01	7	0.75	0.66	0.07
0.01	8	0.67	0.74	0.08
0.01	9	0.57	0.82	0.10
0.01	10	0.47	0.88	0.11
0.01	11	0.36	0.93	0.12
0.01	12	0.25	0.97	0.13
0.01	13	0.13	0.99	0.14
0.02	14	0.01	1.00	0.16
0.02	15	-0.11	0.99	0.17
0.02	16	-0.23	0.97	0.18
0.02	17	-0.34	0.94	0.19
0.02	18	-0.45	0.89	0.20
0.02	19	-0.56	0.83	0.22
0.02	20	-0.65	0.76	0.23
0.02	21	-0.74	0.68	0.24
0.03	22	-0.81	0.58	0.25
0.03	23	-0.88	0.48	0.26
0.03	24	-0.93	0.37	0.28
0.03	25	-0.97	0.26	0.29
0.03	26	-0.99	0.14	0.30

Example 2.1.12: circle in space

Consider a circle parallel to the xy-plane placed within the plane $z = c$:



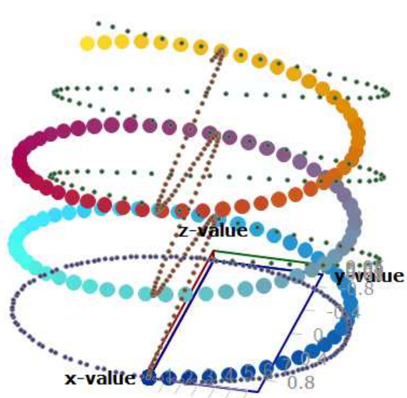


This is its formula:

$$F(t) = (\cos t, \sin t, c) .$$

Example 2.1.13: helix

An ascending spiral is when the rotation in the horizontal plane is combined with a vertical ascend:

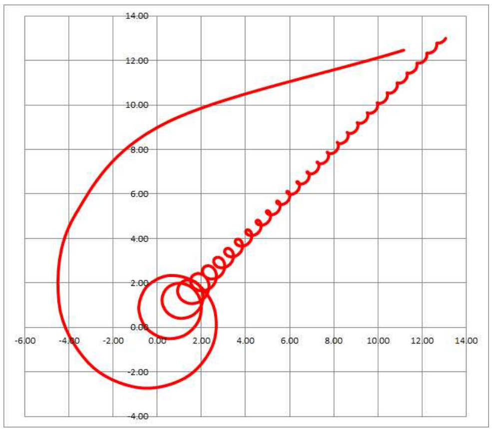


This is its formula:

$$F(t) = (\cos t, \sin t, t) .$$

Exercise 2.1.14

Suggest a parametric representation for the curve below:

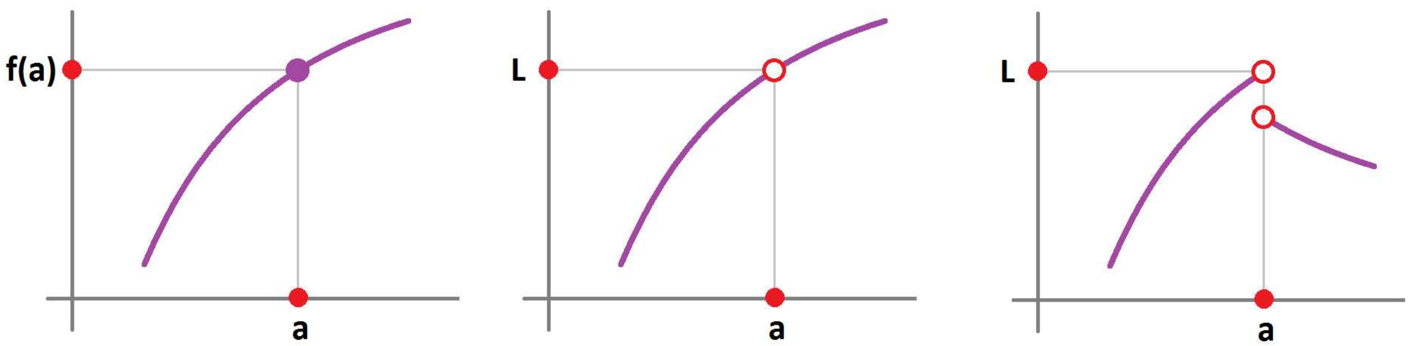


We will revisit every issue about functions we considered in Volumes 2 and 3 as they apply to parametric curves.

But first we consider the issue that lies at the heart of calculus: the rate of change. We take another look at the location-velocity-acceleration connection.

2.2. Limits

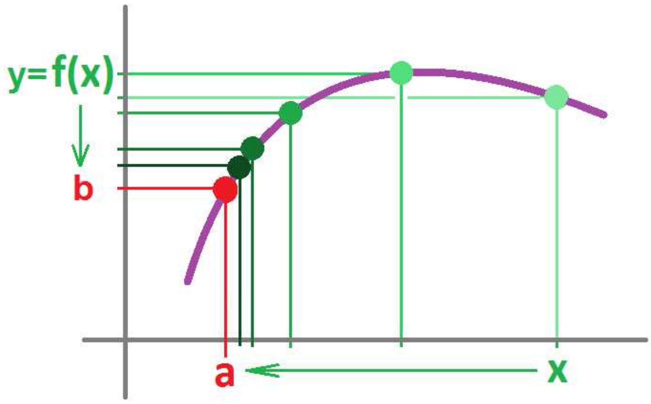
We have always assumed that to get from point A to point B, we have to visit every location between A and B.



If there is a jump in the path of the parametric curve, it can't represent motion! Then the question becomes the one about the integrity of the path: *is there a break or a cut?* Thus, we want to understand what is happening to $x = f(t),\ y = g(t)$ when t is in the vicinity of a chosen input value $t = s$.

This is how we handled the problem in dimension 1. For the curve represented by the graph of $y = f(x)$ we say that $f(x)$ approached b as x approaches a:

$$f(x) \to b \ \text{ as } x \to a\,.$$



The picture can also be described as follows:

- x is approaching a, and
- y is approaching b.

This coordinatewise thinking isn't good enough for us anymore and that is why we rephrase one more time:

- a point (x, y) on the curve is approaching point (a, b).

This way to express the idea of limit is fully applicable to parametric curves! The difference this time is that y doesn't depend on x anymore; instead, $x = x(t)$ and $y = y(t)$ depend on t. This is what we have:

$$(x, y) \to (a, b) \quad \text{as } t \to s\,,$$

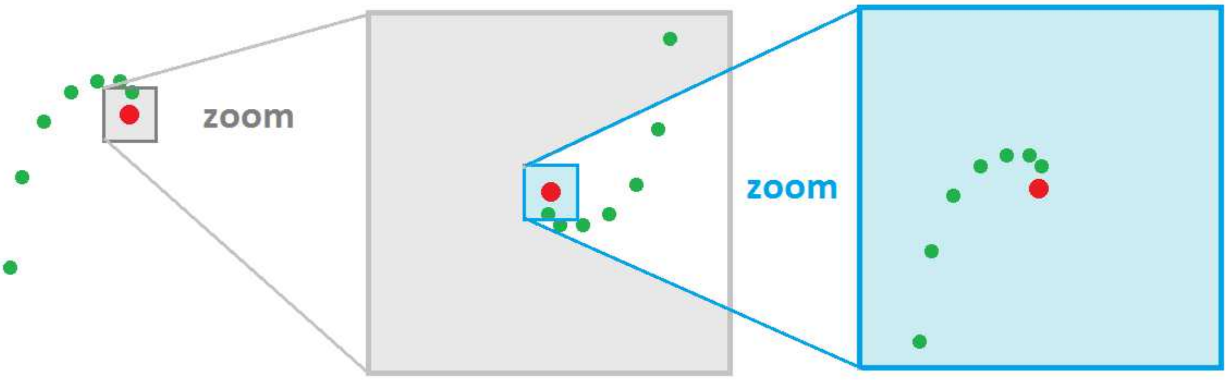
for some real number s.

More generally, we a parametric curve is a function defined on a closed interval in $\mathbf{R}$ with values in $\mathbf{R}^n$ (treated as either points or vectors):

$$X(t) \to A \quad \text{as } t \to s\,.$$

Only the former part needs to be explained.

How do we know that points $X(t)$ are approaching another point A? This means that $X(t)$ is getting closer and closer to A or, more precisely, *the distance from* $X(t)$ is getting smaller and smaller and, in fact, approaching zero.



Definition 2.2.1: limit of parametric curve

The *limit of a parametric curve* $X = X(t)$ in $\mathbf{R}^n$ at $t = s$ is defined to be such a point A in $\mathbf{R}^n$ that

$$d(X(t), A) \to 0 \quad \text{as } t \to s\,.$$

For vectors, the analogous definition is:

$$||X(t) - A|| \to 0 \quad \text{as } t \to s\,.$$

In either case, we use the notation:

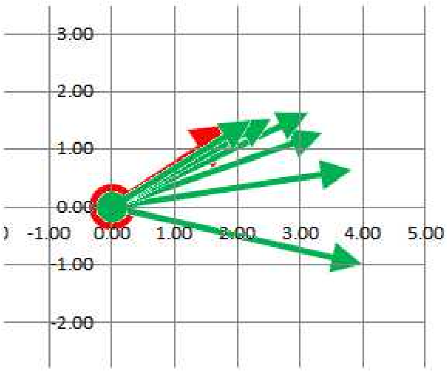
Limit

$$\lim_{t\to s} X(t) = A$$

or

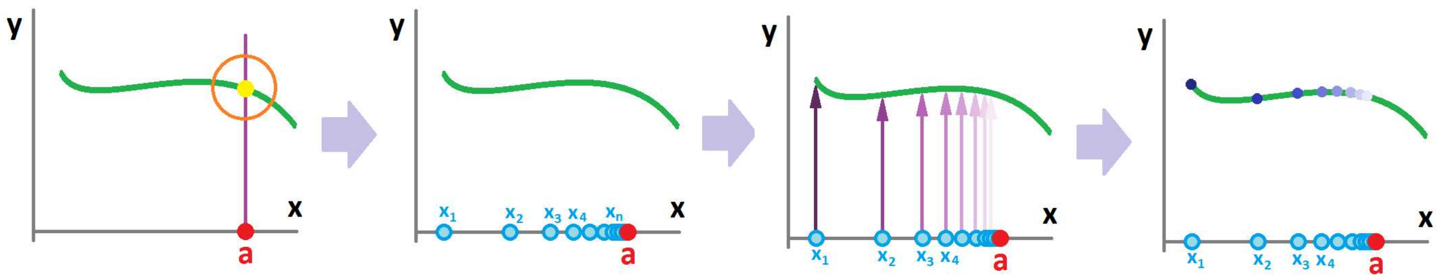
$$X(t) \to A \quad \text{as } t \to s$$

Note how vectors converge in both the magnitude and direction:



We have thus defined a new concept, the limit of a parametric curve, by relying entirely on something familiar: the limit of the usual, *numerical* functions!

Alternatively, we can take a direct route below. We mimic the theory of limits of functions and concentrate on *a sequence converging to the point*.



Example 2.2.2: convergence

Suppose we would like to study this parametric curve around the point $t = 0$:

$$x = \cos x,\ y = x^2 .$$

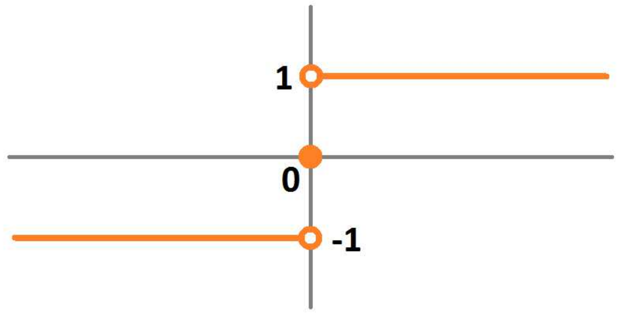
The reciprocal sequence is an appropriate choice:

$$t_n = \frac{1}{n} \to 0 .$$

Recall that the composition of any function and a sequence gives us a new sequence...

Example 2.2.3: the sign

Consider the graph of the function sign around 0. Its standard parametrization is $x = t,\ y = \mathrm{sign}(t)$.



We try $t_n = -1/n$ and $s_n = 1/n$:

$$\lim_{n\to\infty} \operatorname{sign}(-1/n) = -1 \text{ but } \lim_{n\to\infty} \operatorname{sign}(1/n) = 1\,,$$

as we approach 0 from one direction at a time. The limits don't match!

The alternative definition is in the following theorem.

Theorem 2.2.4: Alternative Definition of Limit

The limit of a parametric curve $X = X(t)$ in $\mathbf{R}^n$ at $t = s$ is the limit of these sequences of points in $\mathbf{R}^n$:

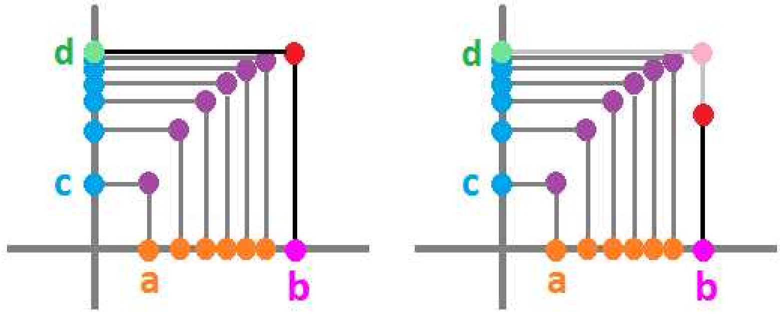
$$\lim_{n\to\infty} X(t_n)$$

considered over all sequences $\{t_n\}$ within the domain of X excluding s that converge to s,

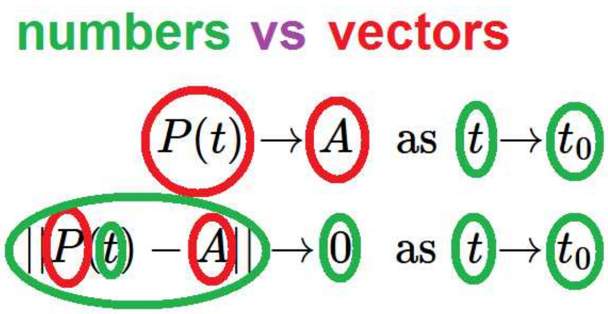
$$s \neq t_n \to s \quad \text{as } n \to \infty\,,$$

when all these limits exist and are equal to each other. Otherwise, the limit does not exist.

This is what happens for $n = 2$:



Let's concentrate on parametric curves as vector-valued function. This is the definition of the limit:



$$P(t) \to A \text{ as } t \to t_0$$

$$||P(t) - A|| \to 0 \text{ as } t \to t_0$$

Since the outputs of these functions are vectors, all the algebraic operations available for vectors are also possible for functions (parametric curves):

- vector addition
- scalar multiplication: the scalar might itself depend on t!

- the dot product: the outcome is a scalar function!

Under these operations, the *limits are preserved.*

Let's consider them one by one.

Theorem 2.2.5: Sum Rule For Limits of Parametric Curves

If the limits at $t = s$ of parametric curves $X = F(t)$ and $X = G(t)$ exist then so does that of their sum, $X = F(t) + G(t)$, and the limit of the sum is equal to the sum of the limits:

$$\boxed{\lim_{t\to s}\big(F(t)+G(t)\big) = \lim_{t\to s} F(t) + \lim_{t\to s} G(x)}$$

Proof.

We can prove the result by using the above theorem; as follows: for any sequence $t \to s$, we have by SR for sequences:

$$\lim_{t\to s}(F(t)+G(t)) = \lim_{n\to\infty}(F(t_n)+G(t_n)) = \lim_{n\to\infty} F(t_n) + \lim_{n\to\infty} G(t_n)\,.$$

Alternatively, we use the definition. If A and B are the two limits, we consider and manipulate the following expression:

$$\begin{array}{rcccccl}
||(F(t)+G(t))-(A+B)|| & = & ||(F(t)-A) & + & (G(t)-B)|| & & \text{Re-arrange terms.}\\
 & \le & ||(F(t)-A)|| & + & ||(G(t)-B)|| & & \text{Use Triangle Inequality.}\\
 & & \downarrow & & \downarrow & & \text{Taking the limits.}\\
 & = & 0 & + & 0 & & \text{Zeros by the definition.}\\
 & = & 0. & & & &
\end{array}$$

By the definition, we conclude that $F(t) + G(t) \to A + B$.

Theorem 2.2.6: Constant Multiple Rule For Limits of Parametric Curves

If the limit at $t = s$ of a parametric curve $X = F(t)$ exists then so does that of its multiple, $X = cF(t)$, and the limit of the multiple is equal to the multiple of the limit:

$$\boxed{\lim_{t\to s} cF(t) = c\cdot\lim_{t\to s} F(t)}$$

Proof.

We can prove the result by using the above theorem just as in the last proof: for any sequence $t \to s$, we have by CMR for sequences:

$$\lim_{t\to s}\big(cF(t)\big) = \lim_{n\to\infty}\big(cF(t_n)\big) = c\lim_{n\to\infty} F(t_n)\,.$$

Alternatively, we use the definition. If A is the limit, we consider and manipulate the following

expression:

$$\begin{array}{rclll}
\big|\big|\big(cF(t)\big)-\big(cA\big)\big|\big| & = & ||c \;\cdot\; (F(t)-A)|| & \text{Factor.}\\
 & = & |c| \;\cdot\; ||(F(t)-A)|| & \text{Use Homogeneity.}\\
 & & || \qquad \downarrow & \text{Taking the limit.}\\
 & = & |c| \;\cdot\; 0 & \text{Zero by the definition.}\\
 & = & 0\,. &
\end{array}$$

By the definition, we conclude that $cF(t)\to cA$.

What if the multiple isn't constant?

Theorem 2.2.7: Variable Multiple Rule

If the limits at $t=s$ of a parametric curve $X=F(t)$ and a function $r=c(t)$ exist then so does that of their product, $X=c(t)\cdot F(t)$, and the limit of the product is equal to the product of the limits:

$$\boxed{\lim_{t\to s}\big(c(t)\cdot F(t)\big)=\Big(\lim_{t\to s}c(t)\Big)\cdot\Big(\lim_{t\to s}F(t)\Big)}$$

Exercise 2.2.8

Modify the proof of the Constant Multiple Rule to prove the last theorem.

Theorem 2.2.9: Dot Product Rule For Limits of Parametric Curves

If the limits at $t=s$ of parametric curves $X=F(t)$ and $X=G(t)$ exist then so does that of their dot product, $r=F(t)\cdot G(t)$, and the limit of the dot product is equal to the dot product of the limits:

$$\boxed{\lim_{t\to s}\big(F(t)\cdot G(t)\big)=\Big(\lim_{t\to s}F(t)\Big)\cdot\Big(\lim_{t\to s}G(t)\Big)}$$

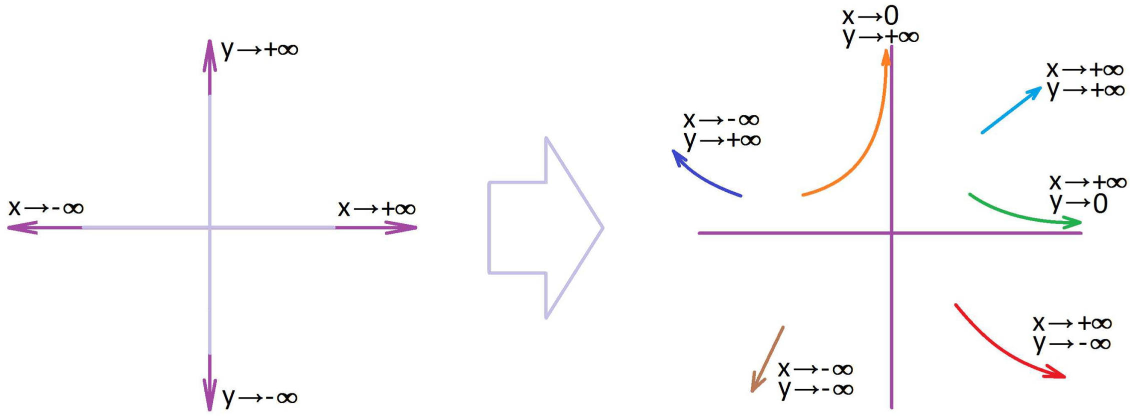
This is a summary of how limits behave with respect to the usual algebraic operations.

Theorem 2.2.10: Algebra of Limits of Parametric Curves

Suppose $F(t)\to A$ and $G(x)\to B$ as $t\to s$. Then

SR:	$F(t)+G(t)\to A+B$	*CMR:*	$cF(t)\to cA$	*for any real c*
DPR:	$F(t)\cdot G(t)\to A\cdot B$	*VMR:*	$c(t)F(t)\to bA$	*when $c(t)\to b$*

Now the *asymptotic behavior* of parametric curves. All graphs are parametric curves, so the latter exhibit just as many types of large-scale behavior as the former.



We will however mention here only the patterns that are coordinate-independent. They are

- convergence to a point
- divergence to infinity
- periodicity

The first one is defined just as the common limit but instead of t approaching some s, $t \to s$, we have t approaching infinity, $t \to \infty$.

Definition 2.2.11: parametric curve approaches point

We say that a parametric curve $X = X(t)$ *approaches point* A as $t \to \pm\infty$ if

$$||X(t) - A|| \to 0 \text{ as } t \to \pm\infty\,.$$

Then we use the notation:

$$X(t) \to A \text{ as } t \to \pm\infty\,,$$

or

$$\lim_{t\to\pm\infty} X(t) = A\,.$$

There are infinitely many directions in $\mathbf{R}^n$, $n > 1$, and that's the reason we consider *only one infinity*: the graph is simply running away from 0.

Definition 2.2.12: parametric curve goes to infinity

We say that a parametric curve $X = X(t)$ *goes to infinity* if

$$||X(t_n)|| \to \infty\,,$$

for any sequence $t_n \to \pm\infty$ as $n \to \infty$. Then we use the notation:

$$X(t) \to \infty \text{ as } t \to \pm\infty\,,$$

or

$$\lim_{t\to\pm\infty} X(t) = \infty\,.$$

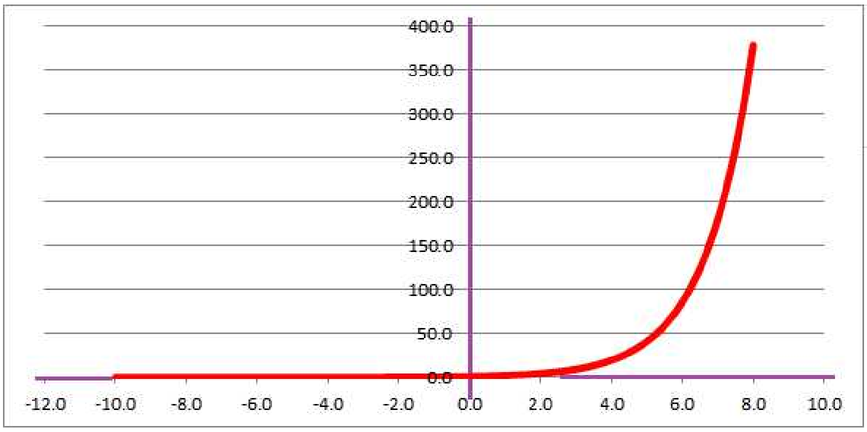
Example 2.2.13: graph

From the familiar facts:

$$\lim_{t\to-\infty} e^t = 0, \quad \lim_{t\to+\infty} e^t = +\infty\,,$$

we deduce that

$$(t, e^t) \to \infty \quad \text{as } t \to \pm\infty \,.$$



Example 2.2.14: Newton's Law of Gravity

We have already seen examples of some of these behaviors exhibited by a planet under the effect of gravity. They may be *bounded* or *unbounded*. Which one we are to observe depends on the initial conditions. Let's make the connection more precise.

2.3. Continuity

A numerical function $y = f(x)$ is called continuous at point $x = a$ if

$$\lim_{x \to a} f(x) = f(a) \,.$$

Thus, the limits of continuous functions can be found by *substitution*.

We approach the issue of continuity of parametric curves in an identical fashion.

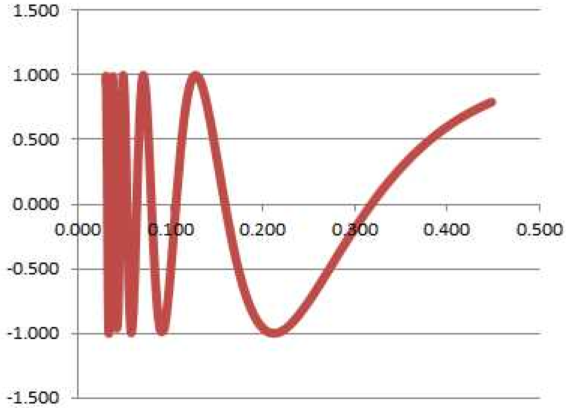
Definition 2.3.1: continuous parametric curve

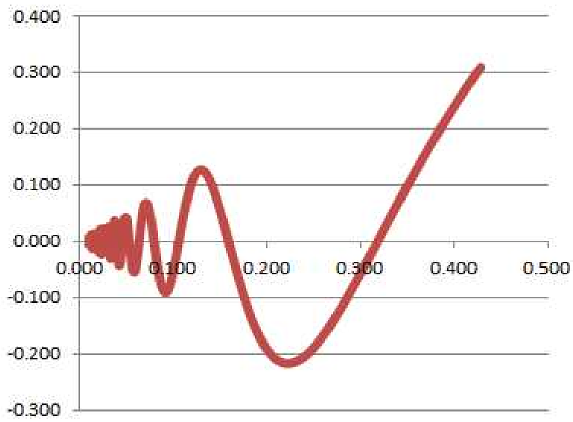
A parametric curve $X = X(t)$ is called *continuous at* $t = s$ if

$$\lim_{t \to s} X(t) = X(s) \,.$$

Example 2.3.2: discontinuity

Examples of continuity and discontinuity come from Volume 2 (Chapter 2DC-2).





Just as before, we can interpret continuity as just another algebraic rule of limits!

Theorem 2.3.3: Substitution Rule

If a parametric curve $X = X(t)$ is continuous at $t = s$ then

$$\boxed{\lim_{t\to s} X(t) = X(s)}$$

A generalization of this rule is related to *compositions*.

Recall that the definition of limit of $X = X(t)$ at $t = s$ can be re-written as an implication:

- If t is approaching s, then $X(t)$ is approaching A.

In other words, we have:

$$t \to s \implies X = X(t) \to A\,.$$

Suppose we have a parametric curve X in $\mathbf{R}^n$. An example of its composition with a function z of n variables was considered in Chapter 1:

$$(z \circ X)(t) = z(X(t))\,.$$

As continuity of functions of several variables hasn't been treated yet, we postpone this example until Chapter 3 and instead look at the composition of X with a numerical function $t = t(u)$ (on the other end):

$$(X \circ t)(u) = X(t(u))\,.$$

Thus we have two functions processing three variables:

$$u \mapsto t \mapsto X\,.$$

Furthermore, let's look at the limit of the composition function at some v:

$$u \to v \implies X = (X \circ t)(u) \to A\,,$$

if this A exists. We break this into two. First:

$$u \to v \implies t = t(u) \to s\,.$$

Second, let's suppose that X is continuous at this $t = s$. Then,

$$t \to s \implies X = X(t) \to X(s)\,.$$

Together:

$$u \to v \implies t = t(u) \to s \implies X = X(t) \to X(s)\,.$$

Theorem 2.3.4: Composition Rule

Suppose $t = t(u)$ is a numerical function with:

$$\lim_{u\to v} t(u) = s\,.$$

Suppose a parametric curve $X = X(t)$ is continuous at $t = s$. Then,

$$\lim_{u\to v} (X \circ t)(u) = X(s)\,.$$

This is a shortened version:

$$\lim_{u\to v} X(t(u)) = X\left(\lim_{u\to v} t(u)\right)\,.$$

It reveals how a continuous function can be pulled out of a limit. In other words, the limit is, again, computed by *substitution*.

We now use the algebraic properties of limits to show how continuity is preserved under these operations.

Theorem 2.3.5: Continuity and Algebra

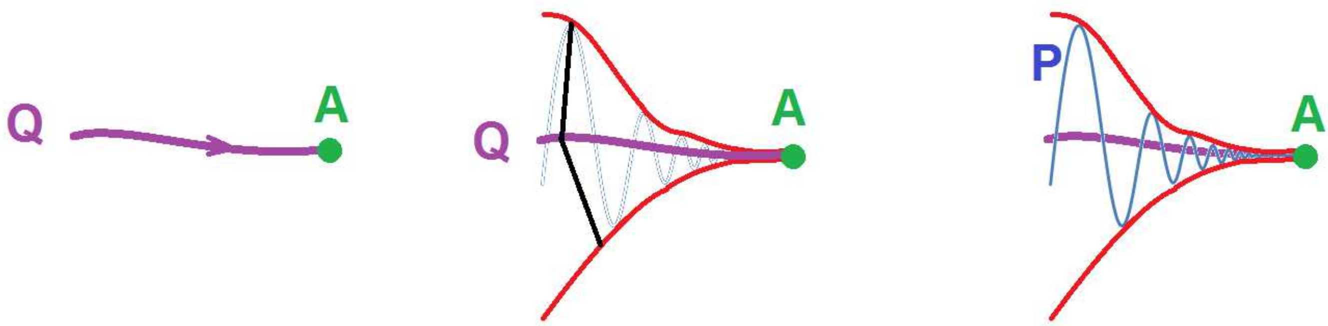
Suppose parametric curves F and G are continuous at $t = s$. Then so are the following:

- *(SR) the parametric curve $F \pm G$,*
- *(CMR) the parametric curve $c \cdot F$, for any real number c,*
- *(VMR) the parametric curve $c \cdot F$, for any continuous numerical function c, and*
- *(DPR) the numerical function $F \cdot G$.*

Some theorems about the behavior of continuous numerical functions have analogues for parametric curves.

First, in the multidimensional spaces there are no "larger" or "smaller" points or vectors. This is why we don't compare functions (or their limits) as easily as $f(x) < g(x)$ anymore. No comparison theorems then. For the same reason we can't easily squeeze a parametric curve between two other parametric curves.

However, we can squeeze a parametric curve with a numerical function used to estimate the distance to another parametric curve. For dimension $n = 2$, this is a narrowing strip:



For dimension $n = 3$, it's a funnel.

Theorem 2.3.6: Squeeze Theorem

Suppose we have first:

$$d(P(t), Q(t)) \le h(t)\,,$$

for all t within some open interval from $t = s$; second:

$$\lim_{t\to s} Q(t) = A\,;$$

and third:

$$\lim_{t\to s} h(t) = 0\,.$$

Then

$$\lim_{t\to s} P(t) = A\,.$$

Proof.

The third condition guarantees that

$$d(P(t), Q(t)) \to 0\,.$$

For the vector-valued functions, the first condition becomes:

$$||P(t) - Q(t)|| \le h(t)\,.$$

The definition of continuity is purely *local*: only the behavior of the function in the, no matter how small, vicinity of the point matters. What can we say about its *global* behavior?

Definition 2.3.7: continuous on interval parametric curve

A parametric curve $P = P(t)$ is called *continuous on interval* I if it is continuous at every s in I.

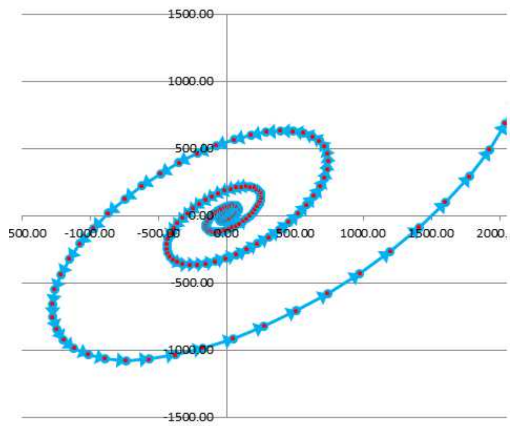
Definition 2.3.8: bounded parametric curve

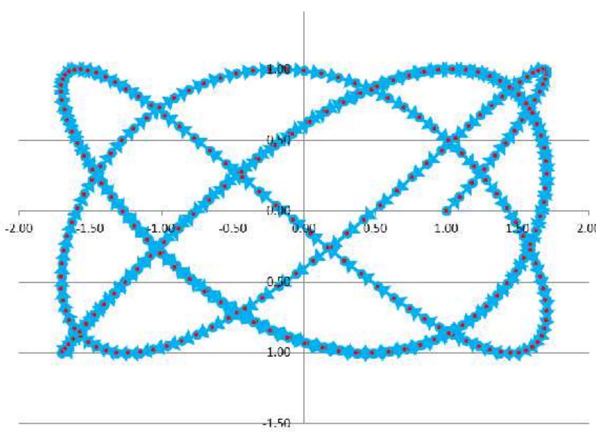
A parametric curve $X = X(t)$ is called *bounded* on an interval $[a,b]$ if there is such a real number m that

$$||X(t)|| \le m$$

for all t in $[a,b]$.

In other words, the *image* of the function is within the sphere of radius m.





Theorem 2.3.9: Boundedness

A continuous on a closed bounded interval parametric curve is bounded on the interval.

Proof.

The proof repeats the one for the 1-dimensional case. Suppose, to the contrary, that X is unbounded on interval $[a,b]$. Then there is a sequence $\{t_n\}$ in $[a,b]$ such that $X(t_n)$ is unbounded. Then, by the *Bolzano-Weierstrass Theorem*, sequence $\{t_n\}$ has a convergent subsequence $\{u_k\}$:

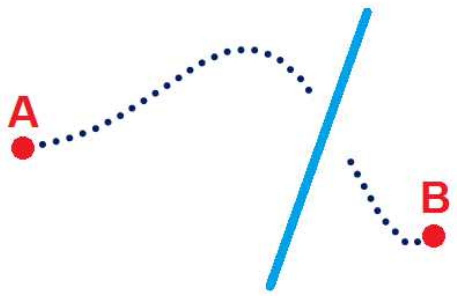
$$u_k \to u\,.$$

This point belongs to $[a, b]$! From the continuity, it follows that

$$X(u_k) \to X(u) \quad \text{or} \quad ||X(u_k) - X(u)|| \to 0\,.$$

This contradicts the fact that $||X(u_k)||$ is a subsequence of a sequence that diverges to infinity.

Our understanding of continuity of numerical functions has been as the property of having no gaps in their graphs. The Intermediate Value Theorem says also that there are no gaps in the ranges either. What about parametric curve? It's *path should have no gaps*. In other words, to get to the other side of the river we have to cross it!



This idea is more precisely expressed by the following.

Theorem 2.3.10: Intermediate Value Theorem

Suppose a parametric curve $X = X(t)$ on the plane is defined and is continuous on interval $[a, b]$. Suppose also that there is a line L such that $X(a)$ and $X(b)$ lie on the different sides of L. Then there is a d in $[a, b]$ such that $X(d)$ lies on L.

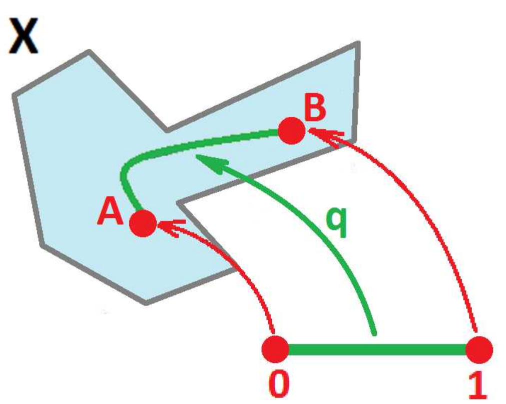
Exercise 2.3.11

Explain the meaning of "lie on the different sides of".

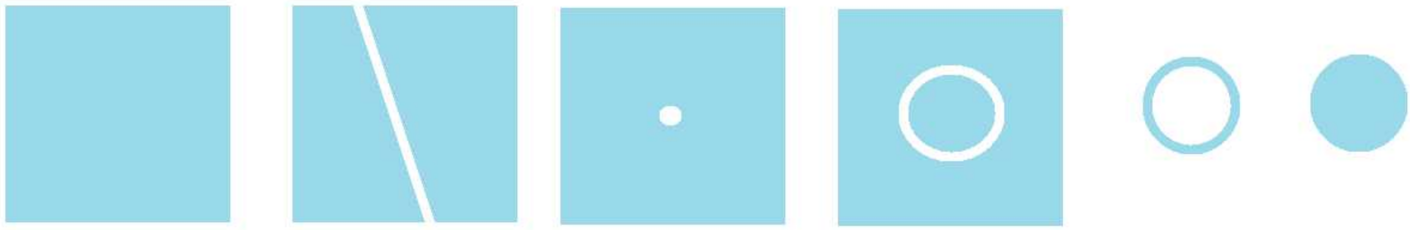
This theorem motivates introducing the following important concept.

Definition 2.3.12: path-connected subset

A subset Q of $\mathbf{R}^n$ is called *path-connected* if any two points in Q can be connected by a path, i.e., if A and B are two points in Q then there is such a continuous parametric curve $X = F(t)$ defined on $[a, b]$ that $F(a) = A$ and $F(b) = B$.



Then the theorem tells us that the plane with a line removed, i.e., $\mathbf{R}^2 \backslash L$ is not path-connected. The plane is, the circle too, a point but not two, a line, a sphere but not two, etc.:



Next, once again, there are no "larger" or "smaller" points or vectors. This is why we can't speak of maximum and minimum points.

Theorem 2.3.13: Extreme Value Theorem

A continuous parametric curve on a bounded closed interval has a point of maximal distance from the origin, i.e., if $X = X(t)$ is continuous on $[a,b]$, then there is c in $[a,b]$ such that

$$||X(c)|| \geq ||X(t)|| \,,$$

for all t in $[a,b]$.

Proof.

It follows from the *Bolzano-Weierstrass Theorem.*

The coordinatewise treatment for parametric curve follows from that for sequences presented in Chapter 1.

Theorem 2.3.14: Coordinatewise Convergence For Parametric Curve

As $t \to s$, we have:

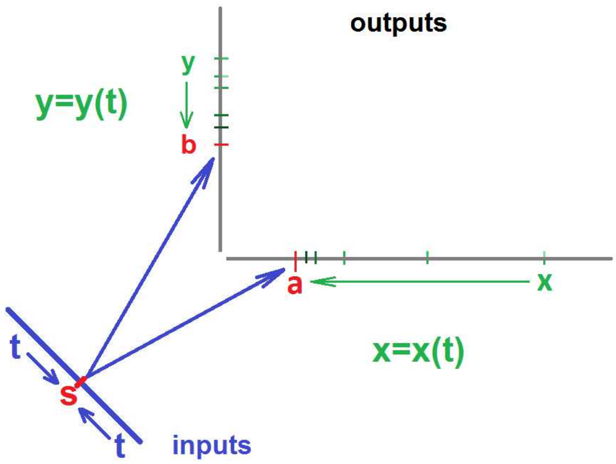
$$X(t) \to A \,,$$

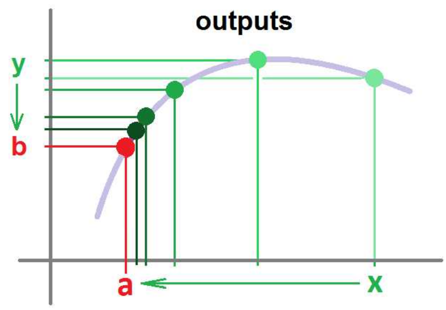
or

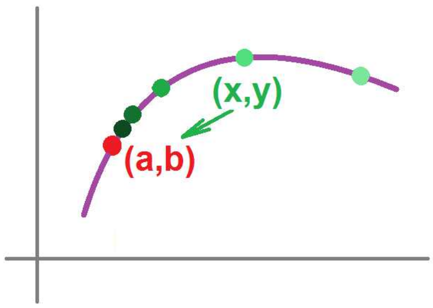
$$X(t) = \big(p_1(t),\ p_2(t),\ \ldots,\ p_n(t)\big) \to A = (a_1,\ a_2,\ \ldots,\ a_n) \,,$$

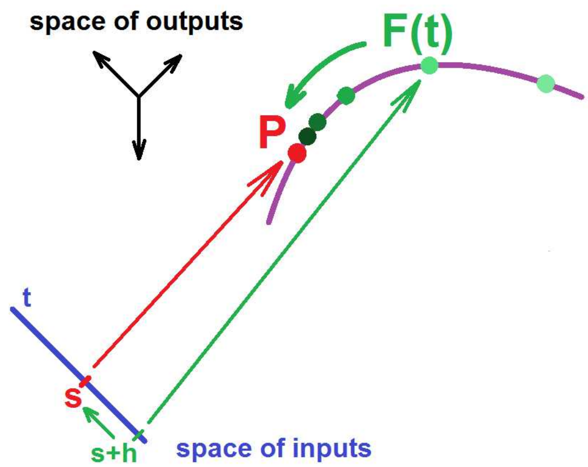
if and only if

$$p_1(t) \to a_1,\ p_2(t) \to a_2,\ \ldots,\ p_n(t) \to a_n \,.$$









Example 2.3.15: a limit

We compute:

$$\begin{aligned}\lim_{t\to\pi/2}(\cos t,\ \sin t) &= \left(\lim_{t\to\pi/2}\cos t,\ \lim_{t\to\pi/2}\sin t\right)\\ &= (\cos\pi/2,\ \sin\pi/2)\\ &= (1,0)\,,\end{aligned}$$

because $\sin t$ and $\cos t$ are continuous.

This theorem also makes it easy to prove the algebraic properties of limits of parametric curves from those for numerical functions. For example, the *Dot Product Rule* is proven below:

$$\begin{aligned}\lim_{t\to s}\big(<x,y>\cdot<u,v>\big) &= \lim_{t\to s}(xu+yv)\\ &= \lim_{t\to s}(xu)+\lim_{t\to s}(yv)\\ &= \lim_{t\to s}x\cdot\lim_{t\to s}u+\lim_{t\to s}y\cdot\lim_{t\to s}v\\ &=<\lim_{t\to s}x,\ \lim_{t\to s}y>\cdot<\lim_{t\to s}u,\ \lim_{t\to s}v>\ .\end{aligned}$$

Now continuity. At $t=s$, suppose a parametric curve

$$X(t)=\big(x(t),\ y(t),\ z(t)\big)$$

is continuous at $t=s$. This is equivalent to the following, as $t\to s$, we have:

$$X(t)\to X(s)\iff x(t)\to x(s),\ y(t)\to y(s),\ z(t)\to z(s)\,.$$

The latter is equivalent to the three component functions $x(t),y(t),z(t)$ being continuous at $t=s$.

Theorem 2.3.16: Coordinatewise Continuity Of Parametric Curve

A parametric curve,

$$X(t) = \big(p_1(t),\ ...,\ p_n(t)\big)\,,$$

is continuous at $t = s$ if and only if all its component functions, $p_1(t),\ ...,\ p_n(t)$, are continuous at $t = s$.

Example 2.3.17: circle

How do we know that the circle is a continuous curve? The two functions in its parametrization $\cos t$, $\sin t$ are continuous.

2.4. Location - velocity - acceleration

We continue with some "pre-limit" calculus; we will revisit the location - velocity - acceleration issue in the context of parametric curves; the motion is on the plane.

Suppose we drove around while paying attention both to the clock and to the mileposts. The result is this simple table with *five* columns:

time (hours):	0	2	4	6	8
location (miles):	$(0,0)$	$(60,0)$	$(120,20)$	$(160,60)$	$(160,100)$

So, we drove east and then north.

What was the velocity over these *four* periods of time? We estimate it with the familiar difference quotient:

$$\text{average velocity} = \frac{\text{change of location}}{\text{change of time}}\,,$$

except this time the numerator, and the average velocity itself, is a *vector*. These are the computations:

time (hours):	0	2	4	6	8
location (miles):	$(0,0)$	$(60,0)$	$(120,20)$	$(160,60)$	$(160,100)$
velocity (m/h):	$\frac{(60,0)-(0,0)}{2-0} =$	$< 30, 0 >$			
velocity (m/h):		$\frac{(120,20)-(60,0)}{4-2} =$	$< 30, 10 >$		
velocity (m/h):			$\frac{(160,60)-(120,20)}{6-4} =$	$< 20, 20 >$	
velocity (m/h):				$\frac{(160,100)-(160,60)}{8-6} =$	$< 0, 20 >$

The four computed values are the average velocities over the following *intervals* of time: $[0,2]$, $[2,4]$, $[4,6]$, and $[6,8]$, respectively. This is, of course, a *partition*. The summary is in this table:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
velocity (m/h):	$< 30, 0 >$	$< 30, 10 >$	$< 20, 20 >$	$< 0, 20 >$

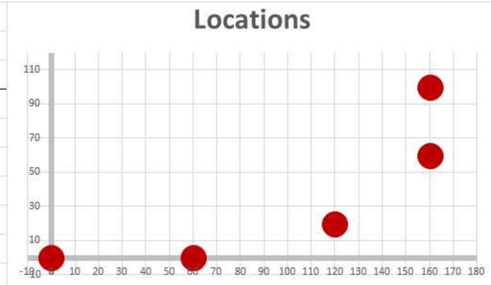
Alternatively, we may choose to assign the four values to the middle points of these intervals, as the *secondary nodes* of the partition, as follows:

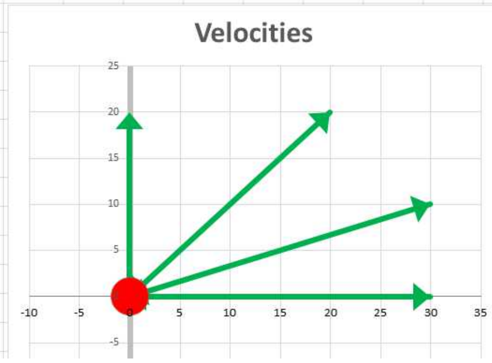
time (hours):	1	3	5	7
velocity (m/h):	$< 30, 0 >$	$< 30, 10 >$	$< 20, 20 >$	$< 0, 20 >$

This amount to simply choosing different secondary nodes in this *partition* of the interval.

This is the summary of what we have found:

time	location		velocity	
	x	y	x	y
h	m	m	m/h	m/h
0	0	0		
1			30	0
2	60	0		
3			30	10
4	120	20		
5			20	20
6	160	60		
7			0	20
8	160	100		





We also compute the *acceleration*. Just as we used the "difference quotient" formula to find the velocity from the location, we now use it to find the acceleration from the velocity, as vectors:

$$\text{average acceleration} = \frac{\text{change of velocity}}{\text{change of time}}\,.$$

We apply this formula to *three* periods of time. These are the computations:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
time (hours):	1	3	5	7
velocity (miles/hour):	$<30,0>$	$<30,10>$	$<20,20>$	$<0,20>$
acceleration (m/h/h):	$\frac{<30,10>-<30,0>}{3-1} =$	$<0,5>$		
acceleration (m/h/h):		$\frac{<20,20>-<30,10>}{5-3} =$	$<-5,5>$	
acceleration (m/h/h):			$\frac{<0,20>-<20,20>}{7-5} =$	$<-10,0>$

The three computed values are the average accelerations over the following *intervals* of time: $[0,4]$, $[2,6]$, and $[4,8]$, respectively. This is the result:

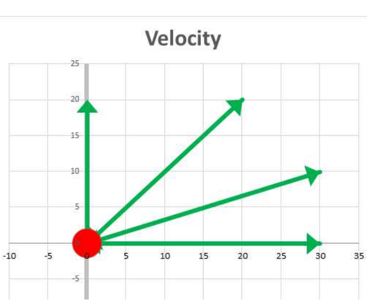
time intervals (hours):	$[0,4]$	$[2,6]$	$[4,8]$
acceleration (m/h/h):	$<0,5>$	$<-5,5>$	$<-10,0>$

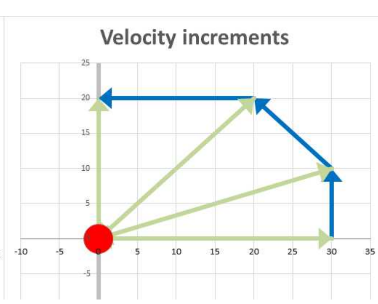
Alternatively, we may choose to assign the three values to the *middle points* of these intervals, as follows:

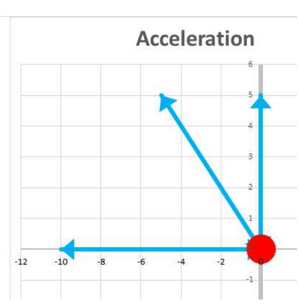
time (hours):	2	4	6
acceleration (m/h/h):	$<0,5>$	$<-5,5>$	$<-10,0>$

This is the summary of what we have found:

time	location		velocity		acceleration	
	x	y	x	y	x	y
h	m	m	m/h	m/h	m/h/h	m/h/h
0	0	0				
1			30	0		
2	60	0			0	5
3			30	10		
4	120	20			-5	5
5			20	20		
6	160	60			-10	0
7			0	20		
8	160	100				







Next we *reverse* the problem: the acceleration is known and the velocity and the location are to be found.

Suppose we start our motion at the initial location $(0,0)$ miles with an initial velocity $<10,0>$ miles per hour and proceed with the acceleration given below:

time intervals (hours):	$[0,4]$	$[2,6]$	$[4,8]$
acceleration (m/h/h):	$<-10,10>$	$<0,10>$	$<10,10>$

We now apply the same formulas as the one we used when the motion was described by a numerical function (or two):

$$\text{next velocity} \;=\; \text{current velocity} \;+\; \text{acceleration} \;\cdot\; \text{time increment}\,,$$

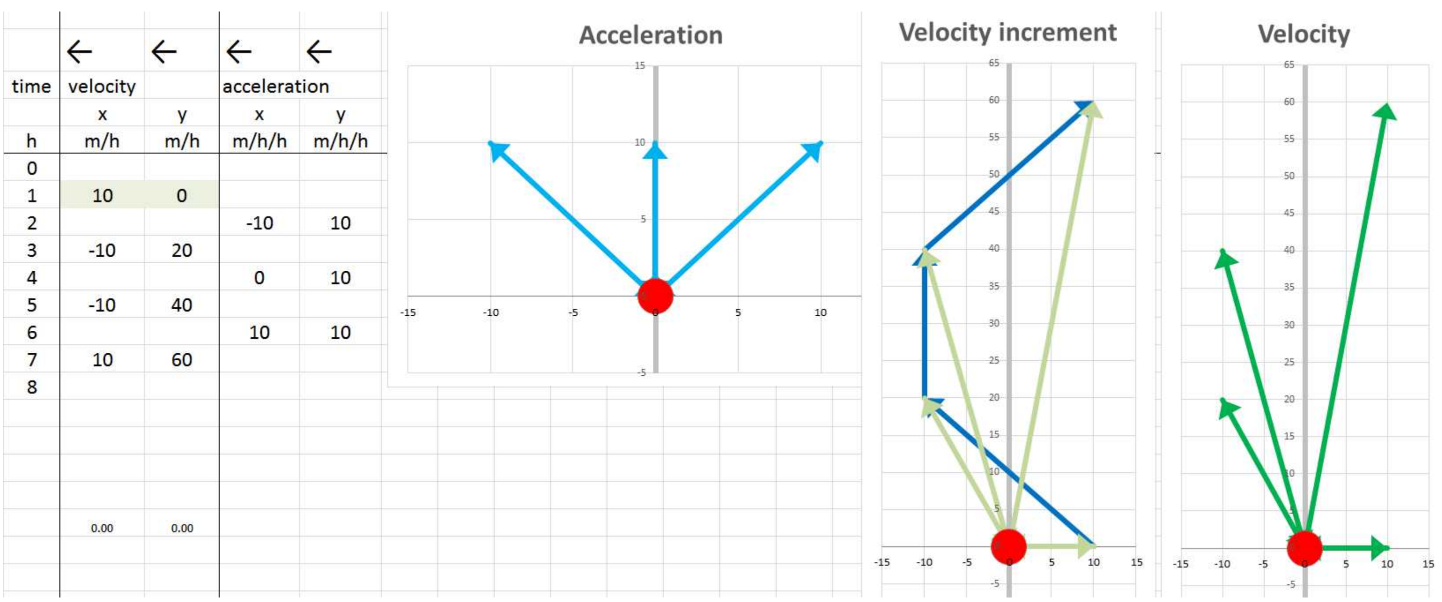
except this time these quantities, except for time, are *vectors*.

We choose the time increment to be 2. These are the computations:

time (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
acceleration (m/h/h):	$<-10,10>$	$<0,10>$	$<10,10>$	
velocity (m/h):	$<10,0>$			
velocity (m/h):	$+2<-10,10>=$	$<-10,20>$		
velocity (m/h):		$+2<0,10>=$	$<-10,40>$	
velocity (m/h):			$+2<10,10>=$	$<10,60>$

The four computed values are the average velocities over the following intervals of time:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
velocity (m/h):	$<10,0>$	$<-10,20>$	$<-10,40>$	$<10,60>$



We compute the position next using the same, vector, formula:

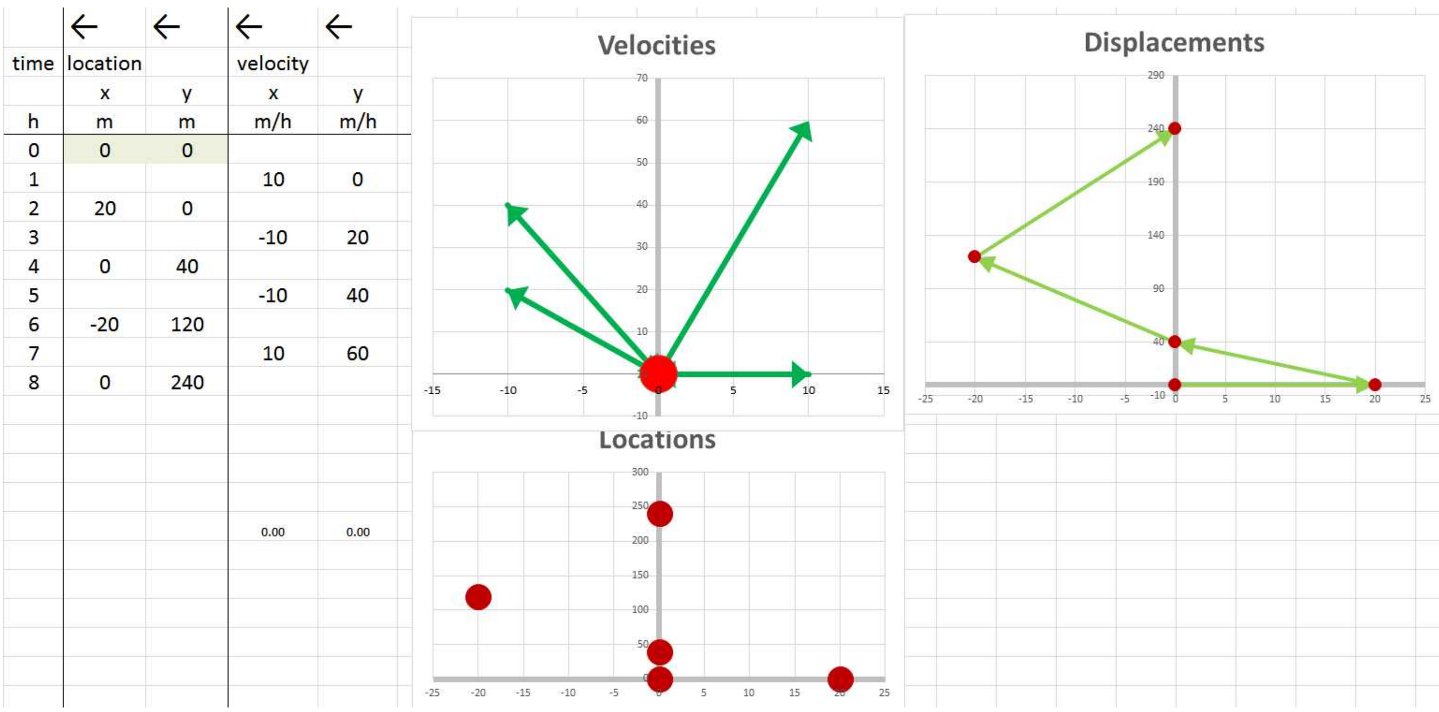
$$\text{next position} \;=\; \text{current position} + \;\text{velocity} \;\cdot\; \text{time increment}\,.$$

These are the results:

time (hours):	0	2	4	6	8
velocity (m/h):	$<10,0>$	$<-10,20>$	$<-10,40>$	$<10,60>$	
position (m):	$(0,0)$				
position (m):	$+2<10,0>=$	$(20,0)$			
position (m):		$+2<-10,20>=$	$(0,40)$		
position (m):			$+2<-10,40>=$	$(-20,120)$	
position (m):				$+2<10,60>=$	$(0,240)$

This is the path:

time (hours):	0	2	4	6	8
location (miles):	$(0,0)$	$(20,0)$	$(0,40)$	$(-20,120)$	$(0,240)$



In summary, we have this collection of sequences for each of the dimensions of the space:

- t_n for the time
- p_n for the position
- v_n for the velocity
- a_n for the acceleration

The last three are linked by these recursive formulas for each component:

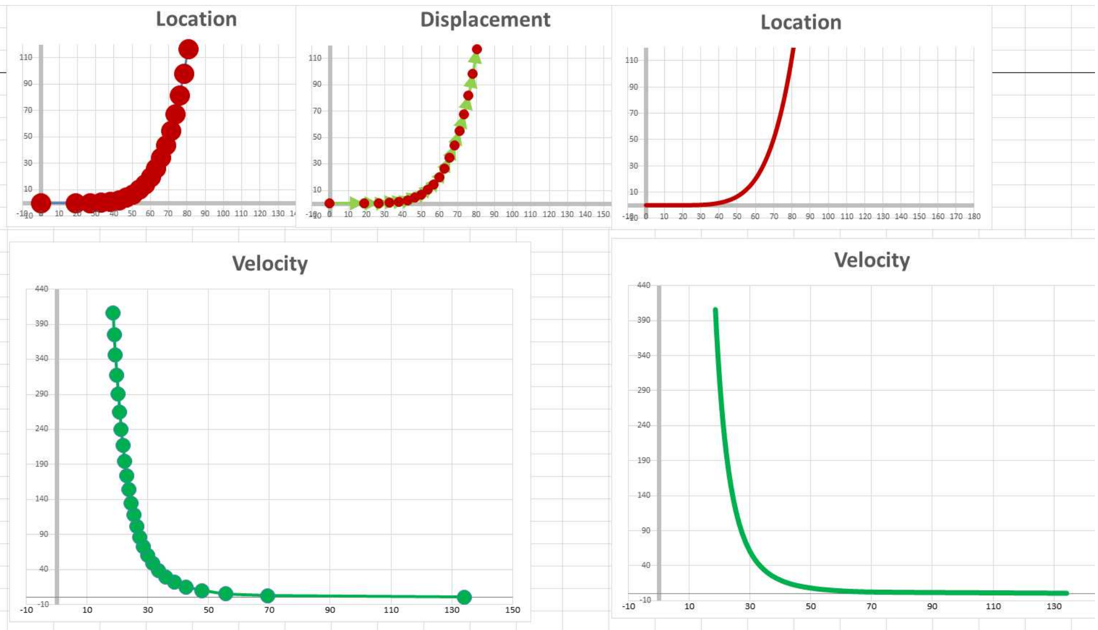
$$v_{n+1} = \frac{p_{n+1} - p_n}{t_{n+1} - t_n} \quad \text{and} \quad a_{n+1} = \frac{v_{n+1} - v_n}{t_{n+1} - t_n}\,.$$

Or, for the inverse problem:

$$v_{n+1} = v_n + a_{n+1}(t_{n+1} - t_n) \quad \text{and} \quad p_{n+1} = p_n + v_{n+1}(t_{n+1} - t_n)\,.$$

With more data, the vectors of the velocities and the accelerations are used to trace out *curves*. For example, this is what computation of the velocity from location looks like:

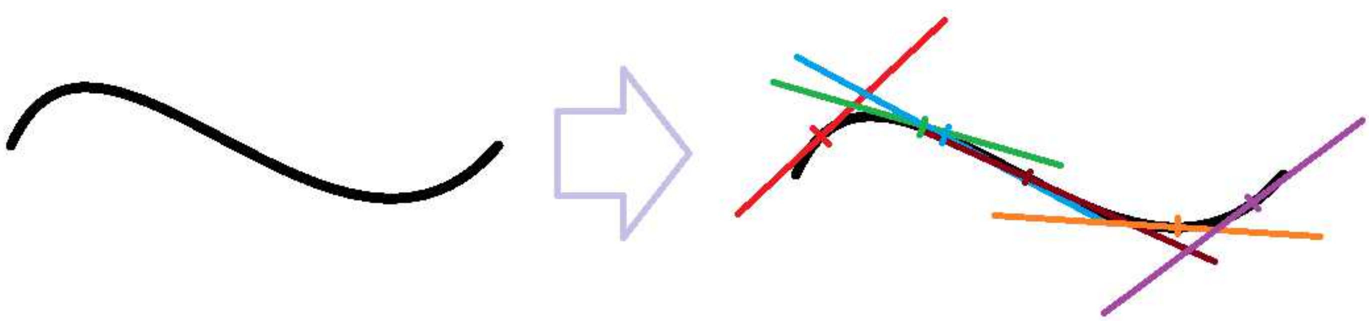
time	location		velocity	
	x	y	x	y
h	m	m	m/h	m/h
0.0	0	0		
0.1	19	0	134	1
0.2	27	0	69	3
0.3	33	1	56	6
0.4	38	1	48	10
0.5	42	3	43	15
0.6	46	4	39	22
0.7	50	7	36	30
0.8	54	10	34	39
0.9	57	15	32	49
1.0	60	20	30	60
1.1	63	27	29	73
1.2	66	35	27	87
1.3	68	44	26	102
1.4	71	55	25	118
1.5	73	68	25	135
1.6	76	82	24	154
1.7	78	98	23	174
1.8	80	117	22	195
1.9	83	137	22	217
2.0	85	160	21	240
2.1	87	185	21	265
2.2	89	213	20	291
2.3	91	243	20	318
2.4	93	276	19	346



Note that the velocity column has one fewer data point and the acceleration one fewer yet. However, when zoomed out, the graphs look like actual continuous curves and give an impression that the three functions have the same domain.

2.5. The change and the rate of change: the difference and the difference quotient

We would like to construct the secant or tangent lines of a parametric curve *at all points* at the same time:



The result will be a function.

Suppose we know only *two* values of a function:

$$F(a) = A \quad \text{and} \quad F(b) = B\,,$$

with $a \neq b$. Then, what can we say about its rate of change? It is the change of X with respect to the change of t. The former is the *difference* of f, denoted as follows:

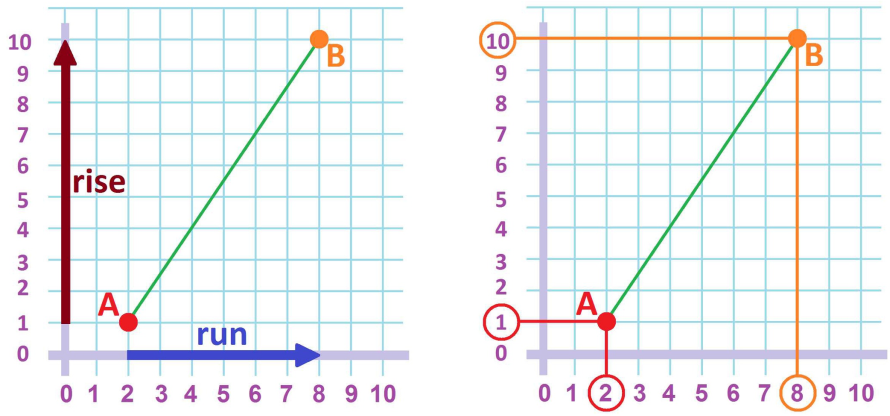
$$\Delta f = B - A = F(b) - F(a)\,.$$

The latter is the *increment* of t, denoted as follows:

$$\Delta t = b - a\,.$$

Their ratio is the rate of change of F, which we will call the *difference quotient* of F, given by:

$$\frac{\Delta F}{\Delta t} = \frac{F(b) - F(a)}{b - a}\,.$$



Let's review the construction of a *augmented partition* X of an interval $[a,b]$. We place the *primary nodes* (or simply nodes) on the interval:

$$a=t_0\le t_1\le t_2\le ...\le t_{n-1}\le t_n=b\,,$$

producing n smaller intervals of possibly different lengths:

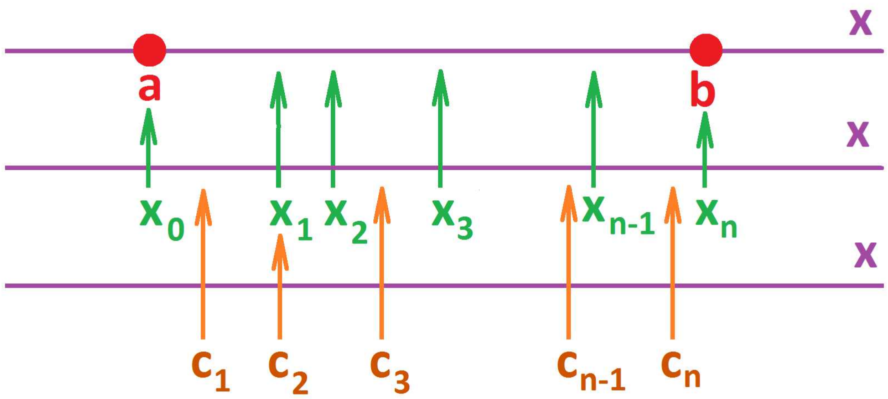
$$[t_0,t_1],\ [t_1,t_2],\ ...,\ [t_{n-1},t_n]\,,$$

with $t_0=a,\ t_n=b$. The lengths of the intervals are the *increments* of t:

$$\Delta t_i=t_i-t_{i-1},\ i=1,2,...,n\,.$$

We are also given the *secondary nodes* of the partition:

$$c_1 \text{ in } [t_0,t_1],\ c_2 \text{ in } [t_1,t_2],\ ...,\ c_n \text{ in } [t_{n-1},t_n]\,.$$



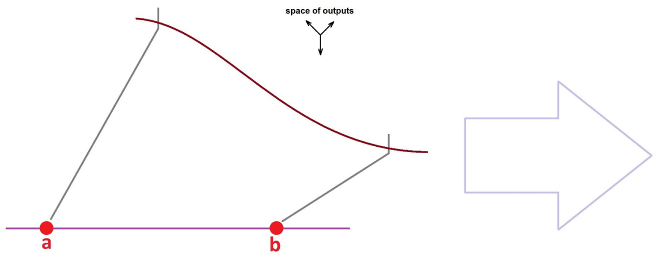
Thus an augmented partition is simply a combination of points:

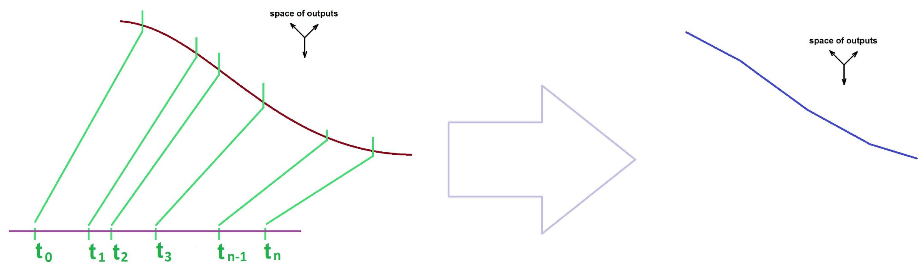
$$a=t_0\le c_1\le t_1\le c_2\le t_2\le ...\le t_{n-1}\le c_n\le t_n=b\,.$$

In the example in the last section we had a similar construction:

- The intervals of the partition were equal in length.
- The secondary nodes were placed at the end of each interval.

Furthermore, we utilize the secondary nodes as the inputs of the new function:





Suppose $X = F(t)$ is defined at the nodes $t_k,\ k = 0, 1, 2, ..., n$, of the partition.

Definition 2.5.1: difference

The *difference* of F is defined at the secondary nodes of the partition by:

$$\boxed{\Delta F\,(c_k) = F(t_k) - F(t_{k-1}),\ k = 1, 2, ..., n}$$

Definition 2.5.2: difference quotient

The *difference quotient* of F is defined at the secondary nodes of the partition by:

$$\boxed{\frac{\Delta F}{\Delta t}(c_k) = \frac{F(t_k) - F(t_{k-1})}{t_k - t_{k-1}} = \frac{\Delta F(c_k)}{\Delta t_k},\ k = 1, 2, ..., n}$$

Note that when secondary nodes aren't provided, we can think of the intervals themselves as the inputs of the difference (and the difference quotient): $c_k = [t_{k-1}, t_k]$. It is then a 1-form, the difference of a 0-form F.

When the context is clear, we use the simplified notation without the subscripts:

Difference and difference quotient

$$\Delta F\,(c) \quad \text{and} \quad \frac{\Delta F}{\Delta t}(c)$$

Example 2.5.3: circle

Let's find the difference quotient of the circle parametrized the usual way:

$$X(t) = (\cos t,\ \sin t)\,.$$

We use the *Trigonometric Formulas* for the difference quotient from Volume 2 (Chapter 2DC-3). Suppose we have a mid-point partition for the interval, say, $[-\pi/2, \pi/2]$, in the t-axis:

- The nodes are $x = a, a + h, ...$ and
- The secondary nodes are $c = a + h/2, ...$.

Then the difference quotients of $\sin x$ and $\cos x$ are given by at c:

$$\frac{\Delta}{\Delta x}(\sin x) = \frac{\sin(h/2)}{h/2} \cdot \quad \cos c; \quad \frac{\Delta}{\Delta x}(\cos x) = -\frac{\sin(h/2)}{h/2} \cdot \quad \sin c\,.$$

Therefore, we can compute coordinatewise:

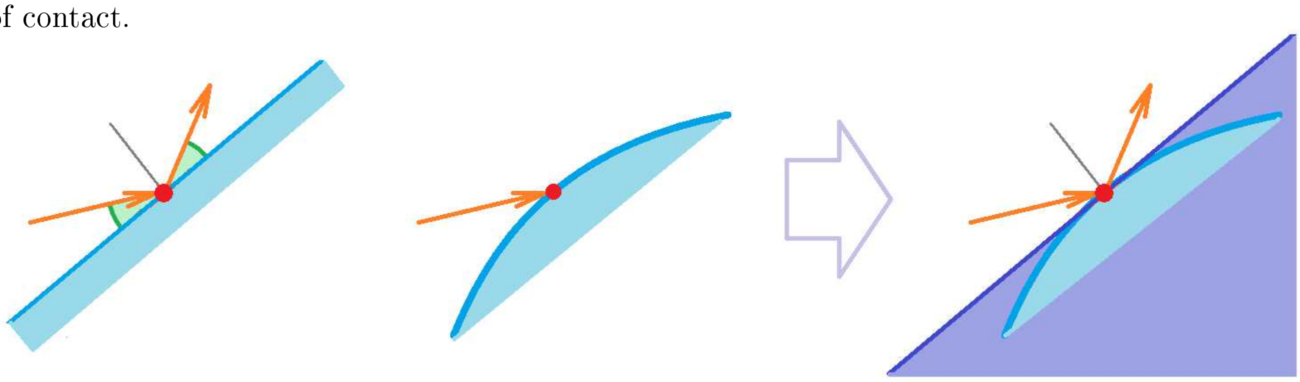
$$\frac{\Delta X}{\Delta t}(c) = \left\langle -\frac{\sin(h/2)}{h/2} \cdot \sin c,\ \frac{\sin(h/2)}{h/2} \cdot \cos c \right\rangle = \frac{\sin(h/2)}{h/2} < -\sin c,\ \cos c > \,.$$

2.6. The instantaneous rate of change: derivative

Let's review a few examples of why we study derivatives. The problems we face are familiar but this time they are treated with *parametric curves*!

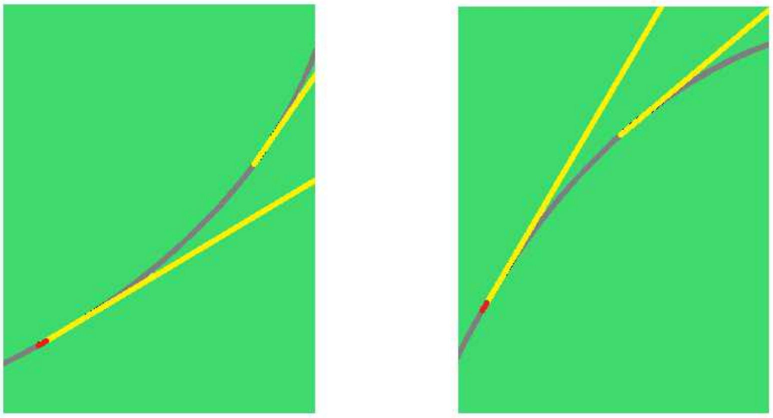
Example 2.6.1: light bounce

Light bounces off a curved mirror as if off a straight mirror that is tangent to the mirror at the point of contact.



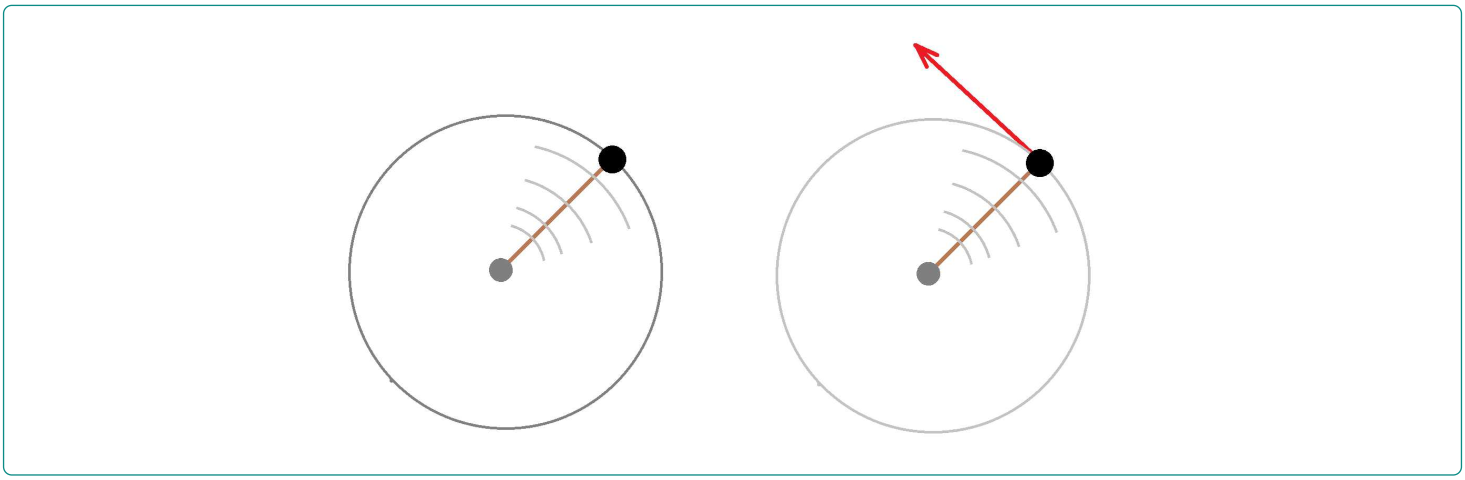
Example 2.6.2: car

The head lights of a car traveling on a curvy road point in the tangent direction to the road at that particular location.

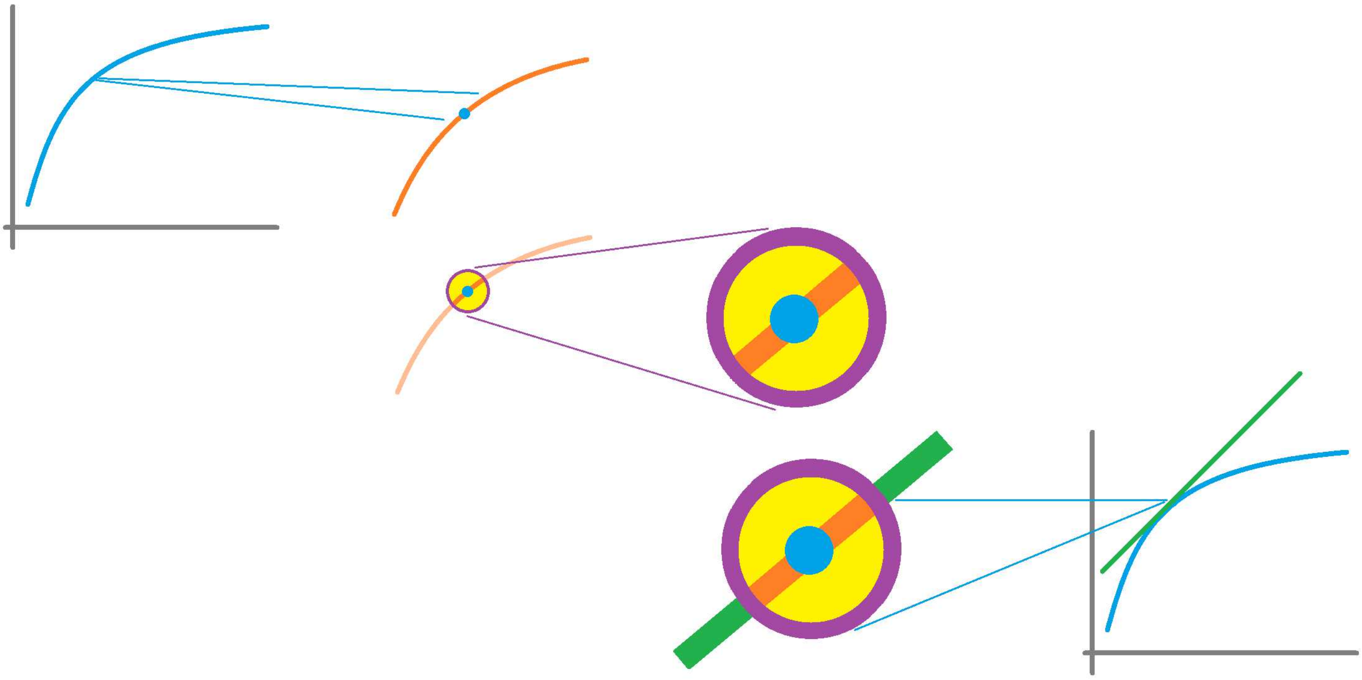


Example 2.6.3: sling

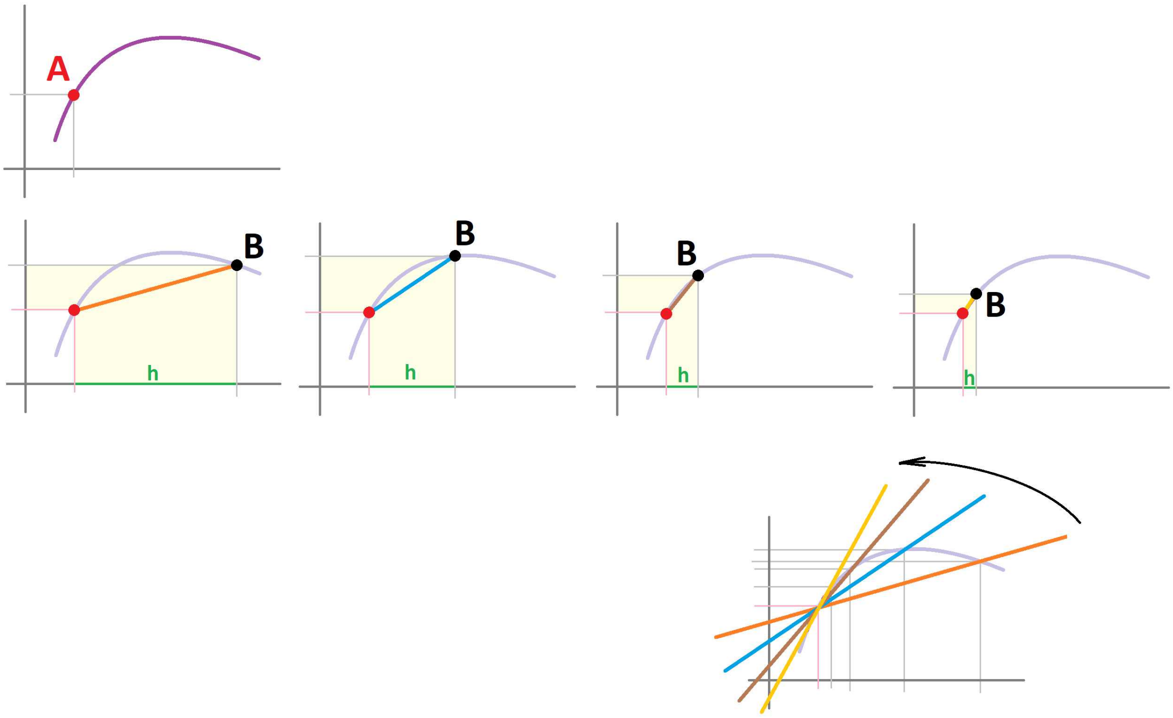
The direction a rock will go when released from a sling is tangent to the circle at the point of release.



Recall how we zoom in on the point and realize that it is virtually a straight line, the *tangent line* at A:

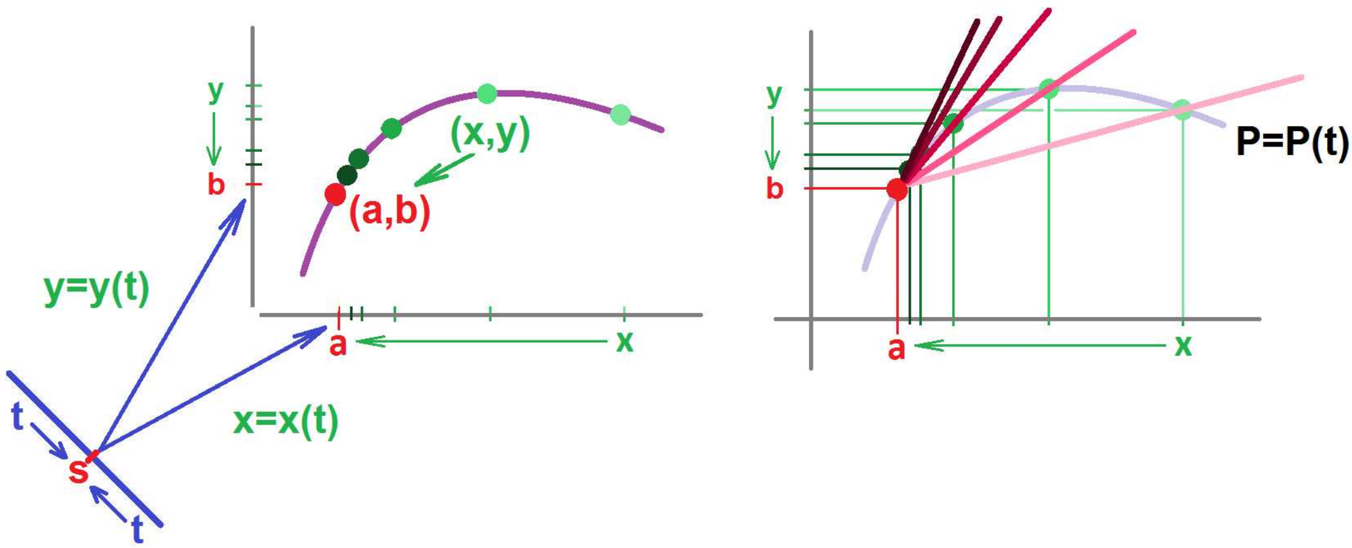


The method of constructing the tangent is to build a sequence of lines that cut through the graph. Each passes through two points, A and some B with $A \neq B$.



What is the difference this time? First we aren't looking for the *slope* of this line anymore but rather its *direction vector*.

Next, with a numerical function, the change of x brings about the change of y and their ratio is the average rate of change and, after the limit, the derivative of the function. With a parametric curve, we do the same for either of its two components: the change of t brings about the change of x and y:



These two ratios are the average rates of change and, after the limit, the two components of the vector of the derivative of the curve.

The coordinate-free interpretation is:

> ▶ The change of t brings about the change of point X and this ratio (of a vector and a number) is the average rate of change and, after the limit, the vector of the derivative of the curve.

This description applies to any dimension:

- t changes from s to $s+h$, and
- X changes from $P=F(s)$ to $Q=F(s+h)$.

Here, h is the increment of the input. It is often denoted as follows:

Increment of the input

$$h = \Delta t$$

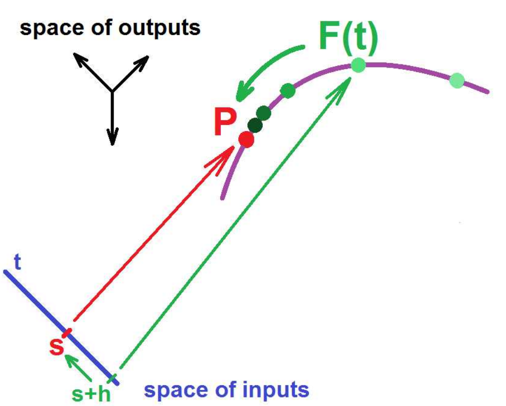
The increment of the output is denoted as follows:

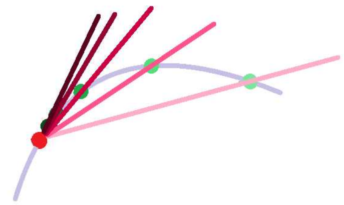
Increment of the output

$$\Delta X$$

Then,

- The change of t is the number h.
- The change of X is the vector XQ.



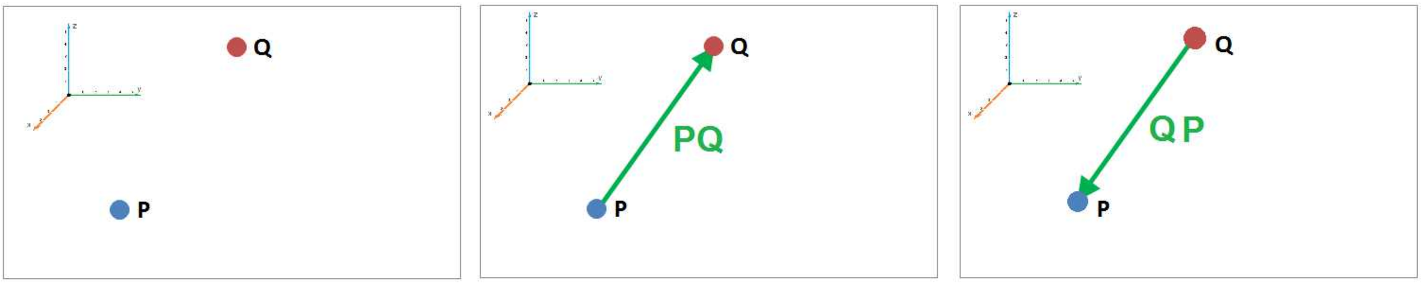


In terms of motion, these are the increment of time and the displacement. We conclude that

▶ The *rate of change* of X with respect to t is vector

$$\frac{1}{h}XQ \quad \text{or} \quad \frac{\Delta X}{\Delta t}\,.$$

It is also known as the *difference quotient* or, in the context of motion, the *average velocity*.



Warning!

When units are involved, as in the case of motion, the vector of the rate of change lies in a space different from the original (and that of the vector of the change of X).

The next step is to make h smaller and smaller. This will make the displacement vector shorter and shorter approaching vector 0 (unless the parametric curve is discontinuous!) but not necessarily the vector of the rate of change. The latter might have a limit!

Definition 2.6.4: derivative

Suppose a parametric curve $X = F(t)$ is defined on an open interval I that contains $t = s$. Then the *derivative* of F at $t = s$ is the following limit, if exists, denoted as follows:

$$\boxed{F'(s) = \lim_{h\to 0} \frac{1}{h}\big(F(s+h) - F(s)\big)}$$

or

$$\boxed{\frac{dF}{dt}(s) = \lim_{h\to 0} \frac{1}{h}\big(F(s+h) - F(s)\big)}$$

When this limit exists, the parametric curve F is called *differentiable* at $t = s$.

Note that the definition is the same if we choose to think of our parametric curve as vector-valued. The result is a new vector-valued function, i.e., a new *parametric curve*!

Example 2.6.5: line

Let's test the definition in a familiar territory. For any two vectors A and B, we have:

$$\begin{aligned}
(At+B)' &= \lim_{h\to 0} \frac{1}{h}\Big[(A(t+h)+B) - (At+B)\Big] \\
&= \lim_{h\to 0} \frac{1}{h}\big[A(t+h) - At\big] \\
&= \lim_{h\to 0} \frac{1}{h}\big[Ah\big] \\
&= \lim_{h\to 0} A \\
&= A\,.
\end{aligned}$$

Then, the derivative of a parametric curve of a line is its direction vector!

More generally, we have a familiar formula for "vector polynomials".

Theorem 2.6.6: Derivative of Polynomials

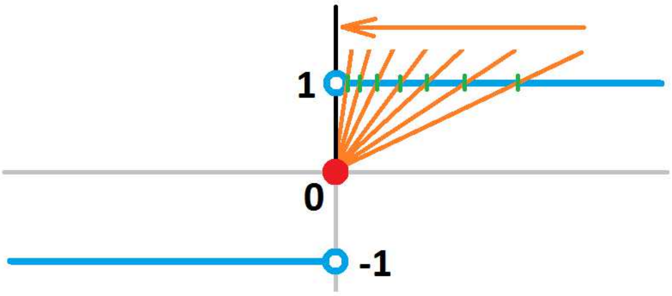
For any vectors $A_m,\ A_{m-1}, ...,\ A_1,\ A_0$ in $\mathbf{R}^n$, we have:

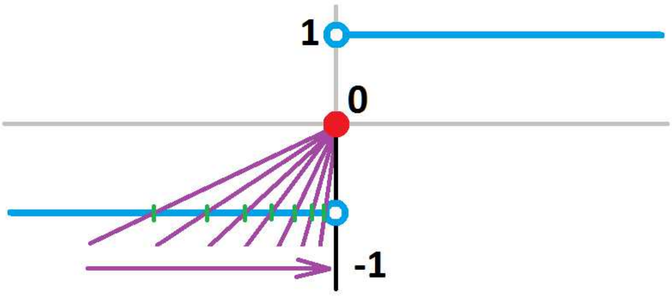
$$\boxed{\Big(A_m t^m + A_{m-1}t^{m-1} + ... + A_1 t + A_0\Big)' = A_m t^{m-1} + A_{m-1}t^{m-2} + ... + A_1}$$

Exercise 2.6.7

Prove the theorem.

When the function is discontinuous, there is a gap in the path. Then the displacements might not converge to 0! In that case, the limit above can't exist.



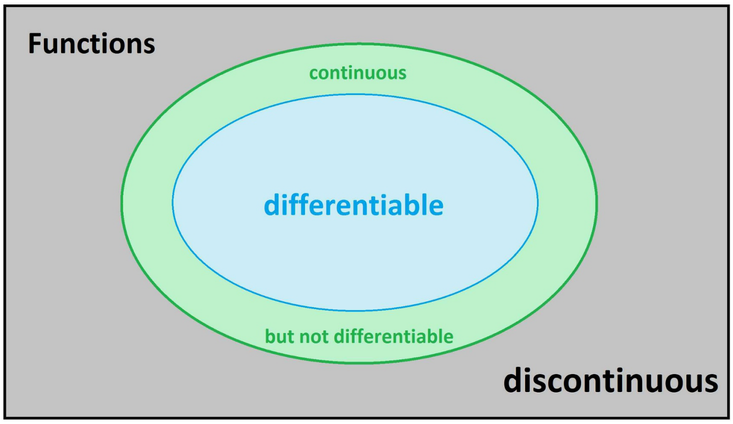


Theorem 2.6.8: Diff => Cont

Every differentiable parametric curve is also continuous.

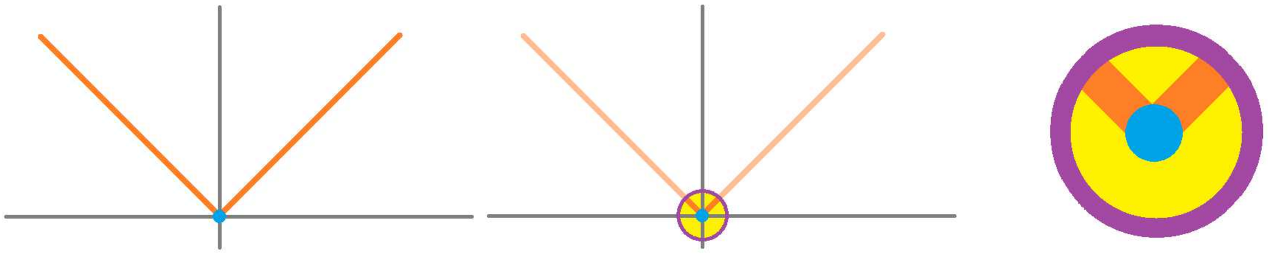
Exercise 2.6.9

Prove the theorem.



Example 2.6.10: non-differentiable

Examples of differentiable and non-differentiable functions come from Volume 2 (Chapter 2DC-3).



If the graph of the absolute value function is parametrized the usual way, $x = t,\ y = |t|$, the resulting parametric curve is continuous but not differentiable.

Exercise 2.6.11

(a) Prove the last statement. (b) Find a differentiable parametrization of $y = |x|$.

2.7. Computing derivatives

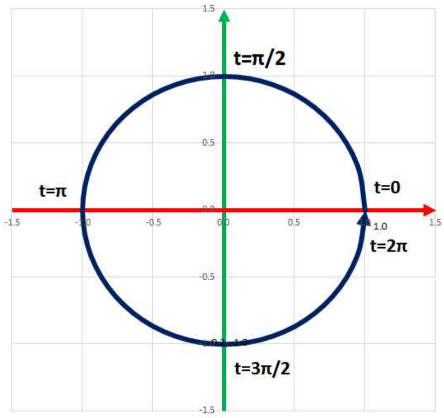
The derivative is a limit and we can use facts about limits we have discussed. We compute derivatives of parametric curves *componentwise*.

Example 2.7.1: circular motion

Let's consider the circle under standard parametrization:

$$F(t) =< \cos t,\ \sin t > \ .$$

We can think of this as if an object is moving around the circle – at a constant *angular velocity*.



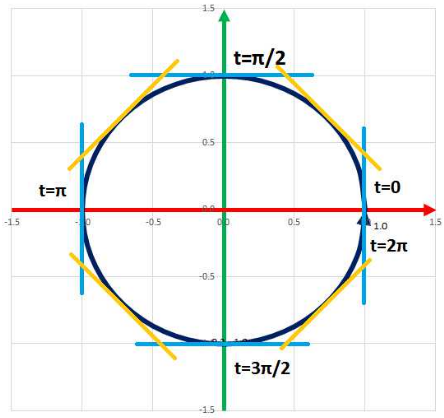
We compute its derivative according to its definition:

$$\begin{array}{lllll}
F'(t) & = \lim_{h\to 0} \frac{1}{h} & (F(t+h) & -F(t)) & \\
 & = \lim_{h\to 0} \frac{1}{h} & \Big(< \cos(t+h),\ \sin(t+h) > & - < \cos t,\ \sin t > \Big) & \text{Substitute.}\\
 & = \lim_{h\to 0} \frac{1}{h} & \Big\langle \cos(t+h)-\cos t, & \sin(t+h)-\sin t \Big\rangle & \text{Vector addition.}\\
 & = \lim_{h\to 0} & \Big\langle \frac{1}{h}(\cos(t+h)-\cos t), & \frac{1}{h}(\sin(t+h)-\sin t) \Big\rangle & \text{Scalar multiplication.}\\
 & = \lim_{h\to 0} & \Big\langle \frac{1}{h}(\cos(t+h)-\cos t), & \frac{1}{h}(\sin(t+h)-\sin t) \Big\rangle & \text{Limit componentwise.}\\
 & = & \Big\langle \lim_{h\to 0} \frac{1}{h}(\cos(t+h)-\cos t), & \lim_{h\to 0} \frac{1}{h}(\sin(t+h)-\sin t) \Big\rangle & \text{The derivatives}\\
 & = & < (\cos t)', & (\sin t)' > & \text{...of the two functions.}\\
 & = & < -\sin t, & \cos t > &
\end{array}$$

Let's confirm the results:

$$\begin{array}{llllll}
F'(0) & =< 0,1 > & \text{vertical,} & F'(\pi/4) & = \Big\langle -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \Big\rangle & \text{diagonal,}\\
F'(\pi/2) & =< -1,0 > & \text{horizontal,} & F'(3\pi/4) & = \Big\langle -\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \Big\rangle & \text{diagonal.}\\
... & & & ... & &
\end{array}$$

These are the direction vectors of the tangent lines:



Further examination reveals that the vector $F'(t)$ of the velocity rotates as t increases. Specifically, $F'(t)$ is perpendicular to $F(t)$:

$$F(t)\cdot F'(t) = < \cos t,\ \sin t > \cdot < -\sin t,\ \cos t > = -\cos t \sin t + \sin t \cos t = 0\,.$$

We confirm what we know from Euclidean geometry: a line tangent to the circle is perpendicular to the corresponding radius.

Further, the object turns the same angle per unit of time and, therefore, covers the same distance on the circle. We confirm that fact by discovering that the object $X = F(t)$ moves at a constant speed:

$$||F'(t)|| = || < -\sin t,\ \cos t > || = \sqrt{(-\sin t)^2 + (\cos t)^2} = 1\ ,$$

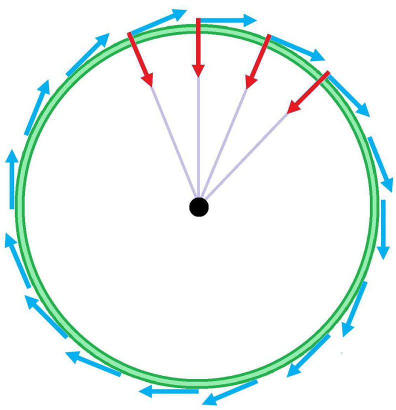
by the *Pythagorean Theorem*. This doesn't mean however that there is no acceleration! The acceleration is what turns the vector F'. We compute it here:

$$F''(t) = (F'(t))' = (< -\sin t,\ \cos t >)' = < -\cos t,\ -\sin t > \ .$$

We can say that F satisfies this *differential equation*:

$$F'' = -F\,.$$

The magnitude of the acceleration, and the force, is constant: $||F''(t)|| = 1$, and it is perpendicular to the velocity F' and, therefore, points at the center of the circle:



Then, there must be something at the origin pulling the object towards it! Indeed, these results match those about planetary motion presented in this chapter (and Chapter 5DE-4).

The following is the general result useful for computations.

Theorem 2.7.2: Componentwise Differentiation

Each component of F' is the derivative of the corresponding component of F.

Exercise 2.7.3

Prove the theorem.

Exercise 2.7.4

Apply the above analysis to a circle of arbitrary radius.

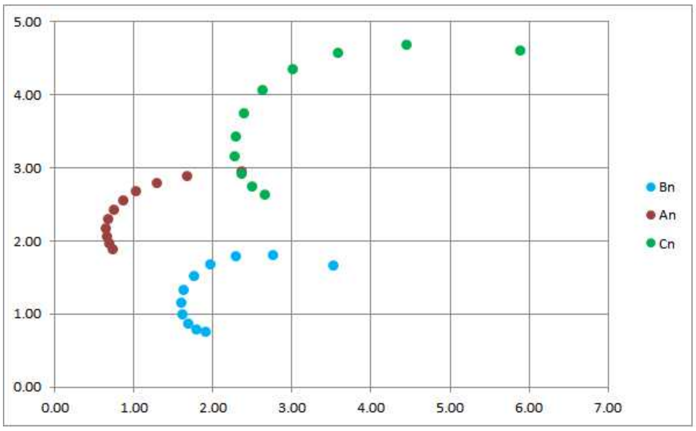
Exercise 2.7.5

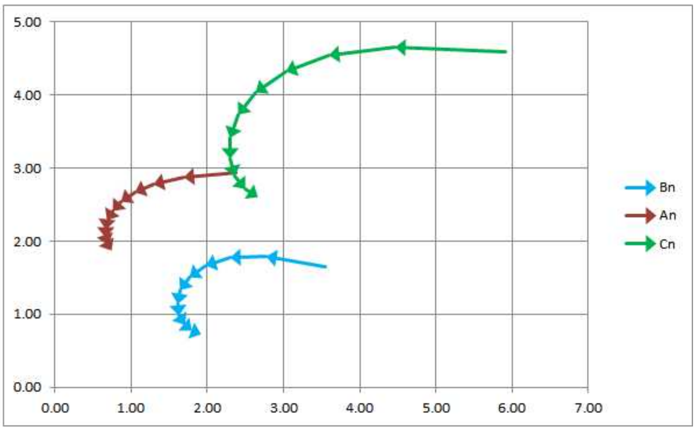
(a) Apply the above analysis to the ellipse. (b) Compare the results to those about planetary motion presented in this chapter.

2.8. Properties of difference quotients and derivatives

More results follow from the rules of limits. All the algebraic rules of differentiation for numerical functions re-appear in this context.

The idea of *addition* is illustrated below:





Here, the vectors are added and so are the vector differences.

Theorem 2.8.1: Sum Rule For Differences

The difference of the sum of two parametric curves is the sum of their differences; i.e., for any two parametric curves $X = F(t)$ and $X = G(t)$ defined at the nodes x and $x + \Delta x$ of a partition, we have the differences defined at the corresponding secondary node c satisfy:

$$\boxed{\Delta\,(F+G)(c) = \Delta F\,(c) + \Delta G\,(c)}$$

Theorem 2.8.2: Sum Rule For Difference Quotients

The difference quotient of the sum of two parametric curves is the sum of their difference quotients; i.e., for any two parametric curves $X = F(t)$ and $X = G(t)$ defined at the nodes x and $x + \Delta x$ of a partition, we have the difference quotients defined at the corresponding secondary node c satisfy:

$$\boxed{\frac{\Delta(F+G)}{\Delta t}(c) = \frac{\Delta F}{\Delta t}(c) + \frac{\Delta G}{\Delta t}(c)}$$

Theorem 2.8.3: Sum Rule For Derivatives

The sum of two functions differentiable at a point is differentiable at that point and its derivative is equal to the sum of their derivatives; i.e., if $X = F(t)$ and $X = G(t)$ are differentiable at $t = s$ parametric curves then so is $X = F(t) + G(t)$ and we have:

$$\boxed{(F+G)'(s) = F'(s) + G'(s)}$$

Proof.

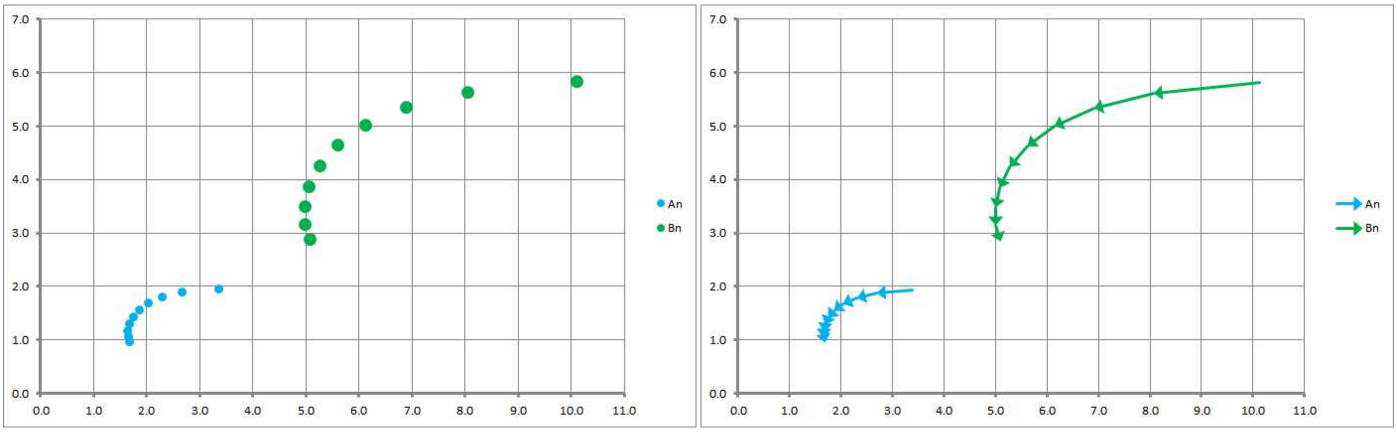
We can use componentwise differentiation and then the *Sum Rule for derivatives of numerical functions*. Alternatively, we use the definition and then the *Sum Rule for limits of parametric curves*.

In terms of motion, if two runners are running starting from a common location, then the vector between them is the sum of their location vectors they have covered.

Exercise 2.8.4

Provide the two proofs.

The idea of *proportion* is illustrated below:



Here, if the location vectors triple then so do their differences.

Theorem 2.8.5: Constant Multiple Rule For Differences

The difference of a multiple of a parametric curve is the multiple of the curve's difference; i.e., for any parametric curve $X = F(t)$ defined at the nodes t and $t + \Delta t$ of the partition, we have the differences defined at the corresponding secondary node c satisfy for any real k:

$$\boxed{\Delta(kf)\,(c) = k\,\Delta f\,(c)}$$

Theorem 2.8.6: Constant Multiple Rule For Difference Quotients

The difference quotient of a multiple of a parametric curve is the multiple of the curve's difference quotient; i.e., for any parametric curve $X = F(t)$ defined at the nodes t and $t+\Delta t$ of the partition, we have the difference quotients defined at the corresponding secondary node c satisfy for any real k:

$$\boxed{\frac{\Delta(kf)}{\Delta t}(c) = k\frac{\Delta f}{\Delta t}(c)}$$

Theorem 2.8.7: Constant Multiple Rule For Derivatives

A multiple of a function differentiable at a point is differentiable at that point and its derivative is equal to the multiple of the function's derivative; i.e., if c is a real number and $X = F(t)$ is a differentiable at $t = s$ parametric curve then

so is $X = cF(t)$ and we have:

$$\boxed{(cF)'(s) = cF'(s)}$$

In terms of motion, if the location vectors are re-scaled, such as from miles to kilometers, then so is the velocity – at the same proportion.

Exercise 2.8.8

Provide the two proofs.

Theorem 2.8.9: Scalar Product Rule For Differences

The difference of a (scalar) product of a numerical functions and a parametric curve is found as a combination of these functions and their differences . In other words, for any function $y = g(t)$ and a parametric curve $X = F(t)$ both defined at the nodes t and $t + \Delta t$ of a partition, we have the differences defined at the corresponding secondary node c satisfy:

$$\boxed{\Delta(gF)\,(c) = g(t + \Delta t)\,\Delta F\,(c) + \Delta g\,(c)\,F(t)}$$

Theorem 2.8.10: Scalar Product Rule For Difference Quotients

The difference quotient of a (scalar) product of a numerical functions and a parametric curve is found as a combination of these functions and their difference quotients. In other words, for any function $y = g(t)$ and a parametric curve $X = F(t)$ both defined at the nodes t and $t+\Delta t$ of a partition, we have and the difference quotients defined at the corresponding secondary node c satisfy:

$$\boxed{\frac{\Delta(gF)}{\Delta t}(c) = g(t + \Delta t)\frac{\Delta F}{\Delta t}(c) + \frac{\Delta g}{\Delta t}(c)F(t)}$$

Theorem 2.8.11: Scalar Product Rule For Derivatives

If $y = g(t)$ is a differentiable at $t = s$ numerical function and $X = F(t)$ is a differentiable at $t = s$ parametric curve then so is $X = g(t)F(t)$ and we have:

$$\boxed{(gF)'(s) = g'(s)F(s) + g(s)F'(s)}$$

Proof.

We start with the difference quotient of the function on the left and then work our way toward the

expression on the right ($h = \Delta t$):

$$\begin{array}{llll}
\multicolumn{3}{l}{\frac{1}{h}\Big(g(s+h)F(s+h) - g(s)F(s)\Big) =} & \\
= \frac{1}{h}\Big(g(s+h)F(s+h) - g(s)F(s+h) & +g(s)F(s+h) - g(s)F(s)\Big) & & \text{Two extra terms.}\\
= \frac{1}{h}\Big(g(s+h)F(s+h) - g(s)F(s+h)\Big) & +\frac{1}{h}\Big(g(s)F(s+h) - g(s)F(s)\Big) & & \text{Combine terms.}\\
= \frac{g(s+h)-g(s)}{h} \cdot F(s+h) & +g(s) \cdot \frac{F(s+h)-F(s)}{h} & & \text{Factor both.}\\
\quad\ \downarrow \qquad\quad \downarrow & \qquad\qquad\quad \downarrow & & h \to 0\\
= \quad g'(s) \quad \cdot F(s) & +g(s) \cdot \quad F'(s), & & \text{Limits according to...}\\
\text{differentiability,} \quad \text{continuity,} & \quad \text{and differentiability,} & &
\end{array}$$

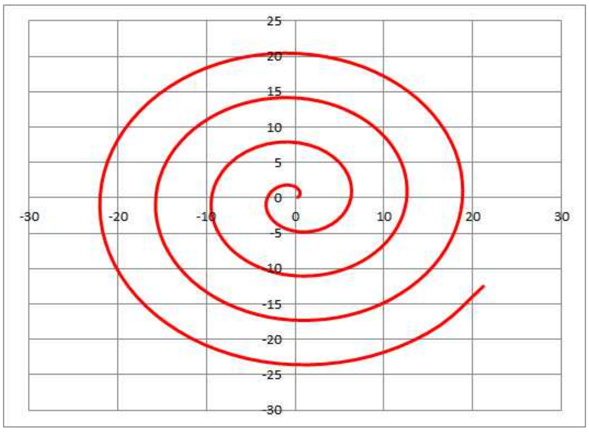
and the *Sum and Product Laws for limits.*

Recall that the rules of vector algebra allow us to carry out algebraic simplification or manipulation with vectors as if they were numbers – as long the expressions themselves make sense. Similarly, the rules of differentiation of parametric curves allow us to carry out differentiation as if these curves were numerical functions – as long the expressions themselves make sense.

Example 2.8.12: spiral

Consider:

$$F(t) = < t\cos t, t\sin t >,\ t \geq 0\,.$$



We apply the *Scalar Product Rule*:

$$\begin{aligned}
F'(t) &= (< t\cos t,\ t\sin t >)'\\
&= (t < \cos t,\ \sin t >)'\\
&= (t)' < \cos t,\ \sin t > + t(< \cos t,\ \sin t >)'\\
&= < \cos t,\ \sin t > + t < -\sin t,\ \cos t > .
\end{aligned}$$

Theorem 2.8.13: Dot Product Rule For Differences

The difference of the dot product of two parametric curves is found as a combination of these functions and their differences. In other words, for two parametric curves $X = F(t)$ and $X = G(t)$ both defined at the nodes t and $t + \Delta t$ of a partition, we have the differences defined at the corresponding secondary node

c satisfy:

$$\boxed{\Delta(F \cdot D)\,(c) = F(t+\Delta t) \cdot \Delta G\,(c) + \Delta F\,(c) \cdot G(t)}$$

Theorem 2.8.14: Dot Product Rule For Difference Quotients

The difference quotient of the dot product of two parametric curves is found as a combination of these functions and their difference quotients. In other words, for two parametric curves $X = F(t)$ and $X = G(t)$ both defined at the nodes t and $t + \Delta t$ of a partition, we have the difference quotients defined at the corresponding secondary node c satisfy:

$$\boxed{\frac{\Delta(F \cdot D)}{\Delta t}(c) = F(t+\Delta t) \cdot \frac{\Delta G}{\Delta t}(c) + \frac{\Delta F}{\Delta t}(c) \cdot G(t)}$$

Theorem 2.8.15: Dot Product Rule For Derivatives

If $X = F(t)$ and $X = G(t)$ are differentiable at $t = s$ parametric curves then $q = F(t) \cdot G(t)$ is a differentiable at $t = s$ numerical function and we have:

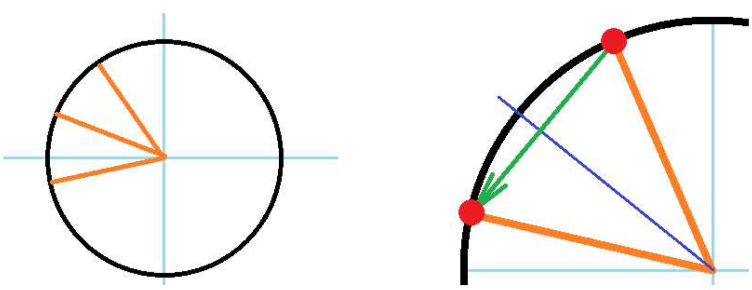
$$\boxed{(F \cdot G)'(s) = F'(s) \cdot G(s) + F(s) \cdot G'(s)}$$

Exercise 2.8.16

Provide the two proofs.

Example 2.8.17: constant elevation

Suppose we are moving along a *circle*. Its center and two consecutive locations form an isosceles. But the median of an isosceles is also its height!

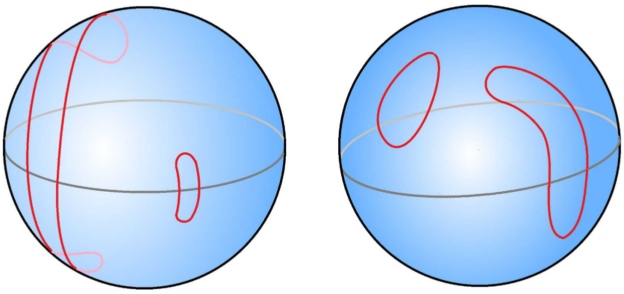


It follows that the displacement is perpendicular to the average of the two consecutive locations. This mimics the familiar fact that the tangent to the circle is perpendicular to its radius. Now, what about a *sphere*?

Suppose that our motion given by $X = F(t)$ is restricted in such a way that the distance to the origin remains the same:

$$||F(t)|| = r\,.$$

In other words, we are moving along the surface of a (hyper)sphere of radius r in $\mathbf{R}^n$, such a circle for $n = 2$ or a sphere for $n = 3$ (such as the surface of the Earth):



It follows

$$\frac{\Delta ||F||^2}{\Delta t} = 0\,.$$

We rewrite as follows:

$$\frac{\Delta\big(F \cdot F\big)}{\Delta t} = 0\,.$$

Then, by the *Dot Product Rule*, we have:

$$F(t+\Delta t) \cdot \frac{\Delta F}{\Delta t}(t) + \frac{\Delta F}{\Delta t}(t) \cdot F(t) = 0\,.$$

Therefore,

$$\frac{\Delta F}{\Delta t}(t) \cdot \left(F(t+\Delta t) + F(t) \right) = 0\,.$$

This means that the difference quotient $\dfrac{\Delta F}{\Delta t}$ is perpendicular to the average of the two consecutive locations.

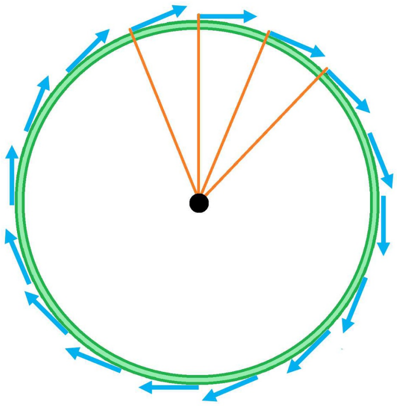
Let's consider the standard parametrization of the circle:

$$F(t) =< \cos t,\ \sin t > \quad \text{and} \quad F'(t) =< -\sin t,\ \cos t > \ .$$

And we know this:

$$F(t) \perp F'(t)\,.$$

What about the sphere?



The argument is identical to that for the difference quotients:

$$\frac{d}{dt}||F||^2 = 0\,.$$

We rewrite as follows:

$$\frac{d}{dt}\left(F \cdot F \right) = 0\,,$$

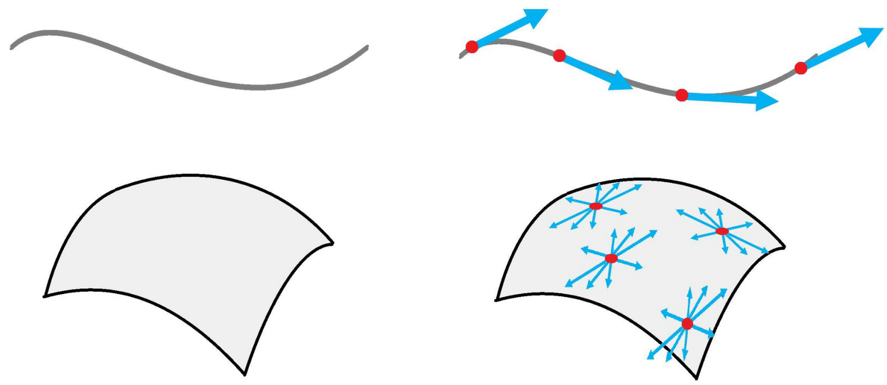
and apply the *Dot Product Rule*:

$$F' \cdot F + F \cdot F' = 0\,.$$

It follows that

$$F' \cdot F = 0\,.$$

This means that the vectors $F(t)$ and $F'(t)$ are perpendicular for every t. The result is to be expected: Our velocity is perpendicular to the radius and, therefore, is *tangential* to the surface of the Earth (pointed neither in nor out):



2.9. Compositions and the Chain Rule

A parametric curve $X = F(t)$ in $\mathbf{R}^n,\ n > 2$,

$$t \mapsto X\,.$$

can be a part of a *composition* in two ways: with a numerical function, $t = g(u)$, before it:

$$u \mapsto t \mapsto X\,.$$

or a function of n variables, $z = q(X)$, after it:

$$t \mapsto X \mapsto z\,.$$

However, the derivative of the latter will only appear in Chapter 3.

Here, we will only concentrate on a simpler case:

$$X = (F \circ g)(u)\,,$$

where g is a numerical function. For example, if t, and u, is time, the function g may be a function of a *change of units* such as from hours to minutes.

Just as the rest of the differentiation rule, this one is also identical to the one for numerical functions: the derivative of the composition is equal to the product of the derivatives.

To understand how the derivatives of these two functions are combined, we start with *linear functions*. In other words, what if we travel along a straight line while also executing a change of units in a linear fashion? After this simple substitution, the derivative is found by direct examination:

		linear function	its derivative	
function of one variable:	$t = g(u)$	$= a + m \cdot (u - v)$	m	in $\mathbf{R}$
	$\circ$			
parametric curve:	$X = F(t)$	$= A + D(t - a)$	D	in $\mathbf{R}^n$
parametric curve:	$F(g(u))$	$= A + D((a + m \cdot (u - v)) - a)$		
		$= A + Dm \cdot (u - v)$	Dm	in $\mathbf{R}^n$

Thus, the derivative of the composition is the *scalar* product of the two derivatives.

We start with the discrete case:

Theorem 2.9.1: Chain Rule For Differences

The difference of the composition of two functions is found as the difference of the latter; i.e., for any function $t = g(u)$ defined at the adjacent nodes u and $u + \Delta u$ of a partition and any parametric curve $X = F(t)$ defined at the adjacent nodes $t = g(u)$ and $t + \Delta t = g(u + \Delta u)$ of a partition, the differences (defined at the secondary nodes v and a within these edges of the two partitions respectively) satisfy:

$$\boxed{\Delta(F \circ g)\,(v) = \Delta F\,(a)}$$

Theorem 2.9.2: Chain Rule For Difference Quotients

The difference quotient of the composition of two functions is found as the product of the two difference quotients; i.e., for any function $t = g(u)$ defined at the adjacent nodes u and $u + \Delta u$ of a partition and any parametric curve $X = F(t)$ defined at the adjacent nodes $t = g(u)$ and $t + \Delta t = g(u + \Delta u)$ of a partition, the difference quotients (defined at the secondary nodes v and a within these edges of the two partitions respectively) satisfy:

$$\boxed{\frac{\Delta(F \circ g)}{\Delta u}(v) = \frac{\Delta F}{\Delta t}(a) \cdot \frac{\Delta g}{\Delta u}(v)}$$

Theorem 2.9.3: Chain Rule For Derivatives

The composition of a function differentiable at a point and a function differentiable at the image of that point is differentiable at that point and its derivative is found as the product of the two derivatives; i.e., if $t = g(u)$ is a differentiable at $u = v$ numerical function and $X = F(t)$ is a differentiable at $t = g(v)$ parametric curve then $X = F(g(u))$ is a differentiable at $u = v$ parametric curve and we have:

$$\boxed{\frac{d(F \circ g)}{dt}(v) = \frac{dF}{dt}(a) \cdot \frac{dg}{du}(v)}$$

On the left, we see simply a new parametric curve, while on the right the parametric curve is multiplied by a variable scalar.

Proof.

It is proven componentwise from the *Chain Rule* for numerical functions.

Example 2.9.4: change of units

An illustration of this conclusion may be a situation when we switch the units of time from minutes to hours,

$$g(u) = u/60,\ g'(u) = 1/60\,,$$

and then realize that the velocity F' in miles per hour will be $1/60$ of the velocity in miles per minute.

Example 2.9.5: accelerated rotation

This may look like speeding up along the circle:

$$G(u) =< \cos u^2,\ \sin u^2 >\,,$$

but can also be seen as motion with accelerated time:

$$G = F \circ g\,,$$

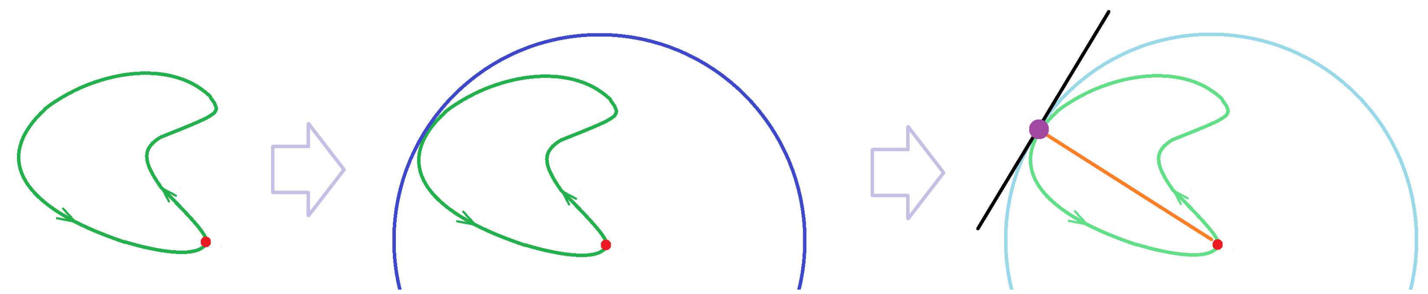
where

$$F(t) =< \cos t,\ \sin t >,\ g(u) = u^2\,.$$

2.10. What the derivative says about the difference quotient: the Mean Value Theorem

We know that if $X = F(t)$ follows a circle or a sphere, the vectors $F(t)$ and $F'(t)$ are perpendicular for every t. What about other curves and surfaces?

In general, this doesn't have to be the case, but one can guess that the trip that requires us to *turn around* will have the direction of the motion perpendicular to the starting point of the trip. We are talking about *round trips*. Indeed, whenever we are the *farthest* from home or any location, we are moving in the direction perpendicular to the direction home, at least for an instant.



Let's make this observation mathematical and turn it into a theorem. We again assume that x is time, limited to interval $[a, b]$.

Theorem 2.10.1: Rolle's Theorem

Suppose a parametric curve $X = F(t)$ satisfies:

- *1. F is continuous on $[a, b]$.*
- *2. F is differentiable on (a, b).*
- *3. $F(a) = F(b) = 0$.*

Then $F'(c) \cdot F(c) = 0$ for some c in (a, b).

Proof.

We will be looking for the farthest location. To find it, we follow the idea of the construction for the circle: consider the distance from the origin, or better the square of the distance, as a function of t. We define a new *numerical function*:

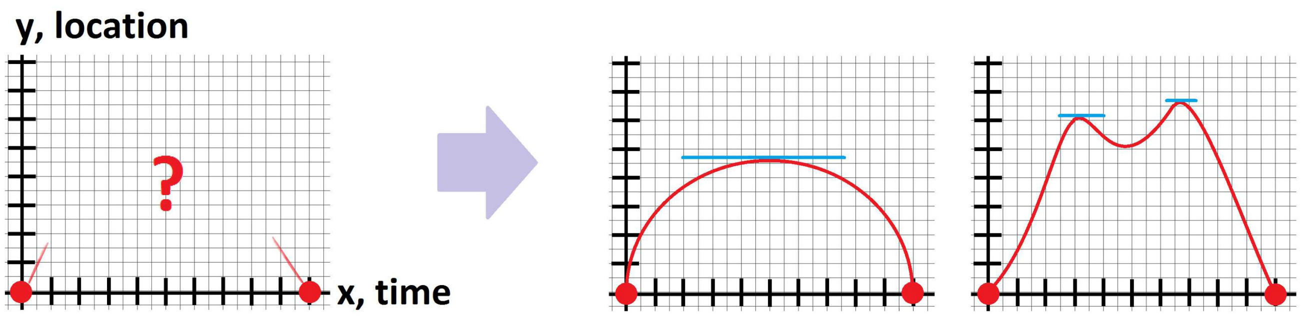
$$g(t) = ||F(t)||^2\,.$$

It is continuous on $[a, b]$ and is differentiable on (a, b). Also

$$g(a) = g(b) = 0\,.$$

Then, by the original *Rolle's Theorem* for numerical functions in Volume 2 (Chapter 2DC-5), there is such a c in (a, b) that

$$g'(c) = 0\,.$$



Then, we have at c:

$$\frac{d}{dt}||F||^2 = 0 \implies \frac{d}{dt}\big(F \cdot F\big) = 0 \implies F' \cdot F + F \cdot F' = 0\,,$$

by the *Dot Product Rule*. It follows that

$$F' \cdot F = 0\,.$$

The condition of the theorem simply states that the difference quotient is zero,

$$\frac{\Delta F}{\Delta t} = 0\,,$$

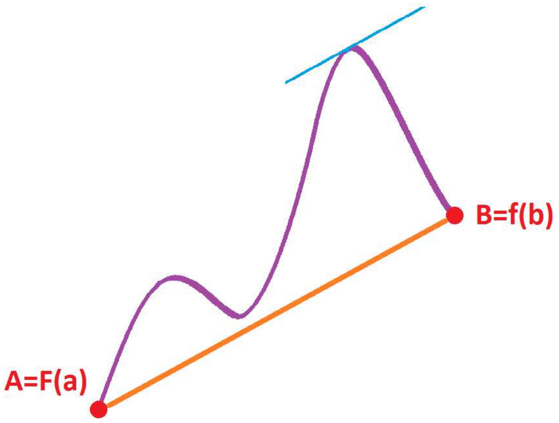
when the partition of $[a, b]$ is trivial: $n = 1,\ x_0 = a,\ x_1 = b$.

The proof indicates that if the average rate of change is zero then the instantaneous rate of change of the distance to the beginning is zero too at some point.

Exercise 2.10.2

Derive the original Rolle's Theorem from this theorem.

Furthermore, what if this isn't a round trip?



The picture suggests what happens to the entities we considered in Rolle's theorem: there is now a line that connects the end-points of the curve and this line is *parallel* to the tangent at some point!

Theorem 2.10.3: Mean Value Theorem

Suppose

1. *F is continuous on $[a, b]$,*
2. *F is differentiable on (a, b).*

Then

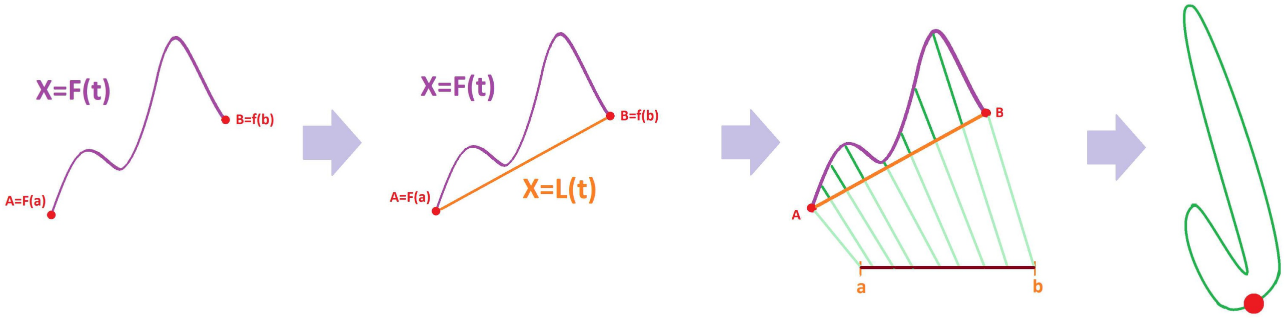
$$\frac{F(b) - F(a)}{b - a} = F'(c)\,,$$

for some c in (a, b).

Proof.

The proof repeats the proof of the original Mean Value Theorem for numerical functions in Volume 2 (Chapter 2DC-5). Let's rename F in *Rolle's Theorem* as H to use it later. Then its conditions take this form:

1. H continuous on $[a,b]$.
2. H is differentiable on $[a,b]$.
3. $H(a)=H(b)$.



Suppose $X=L(t)$ is a parametric curve represented by the straight line from $F(a)$ to $F(b)$ with $L(a)=F(a)$ and $L(b)=F(b)$. Then its derivative is simply its direction vector:

$$L'(t)=\frac{F(b)-F(a)}{b-a}\,.$$

The key step is to define h as the difference of these two parametric curves:

$$H(x)=F(x)-L(x)\,.$$

We now verify the conditions above.

1. H is continuous on $[a,b]$ as the difference of the two continuous functions (SR).

2. H is differentiable on (a,b) as the difference of the two differentiable functions (SR):

$$H'(x)=F'(x)-\frac{F(b)-F(a)}{b-a}\,.$$

3. We also have:

$$F(a)=L(a),\ F(b)=L(b)\implies H(a)=0,\ H(b)=0\implies H(a)=H(b)\,.$$

Thus, H satisfies the conditions of *Rolle's Theorem*. Therefore, the conclusion is satisfied too:

$$H'(c)=0$$

for some c in (a,b); i.e.,

$$F'(c)-\frac{F(b)-F(a)}{b-a}=0\,.$$

Geometrically, c is found by shifting the displacement vector until it touches the graph:

The derivative is defined as the limit of the difference quotient; now we also have a back link. Indeed, what

if we take the partition of $[a,b]$ to trivial: $n=1,\ x_0=a,\ x_1=b$? The theorem simply states that the two are equal, provided we choose the right point where to sample the derivative. For an arbitrary partition, we carry out the construction for each interval of the partition, as follows.

Corollary 2.10.4: Mean Value Theorem For Partition

Suppose:

1. *F is continuous on $[a,b]$.*
2. *F is differentiable on (a,b).*

Then for any partition of the interval $[a,b]$, there are such secondary nodes $c_1,...,c_n$ that

$$\frac{\Delta F}{\Delta x}(c_k)=F'(c_k),\ k=1,2,...,n\,.$$

In other words, the directions of the lines connecting the points of the path of F at the nodes of the partition are equal to the values of the derivative at the secondary nodes between them:

The Mean Value Theorem will help us to derive facts about the function from the facts about its derivative. Before the Mean Value Theorem, we have only been able to find facts about the derivative from the facts about the parametric curve. This is a short list of familiar facts:

info about F		info about F'
F is constant	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	F' is zero
F is linear	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	F' is constant
F is quadratic	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	F' is linear

Are these arrows reversible? If the derivative of the function is zero, does it mean that the function is constant?

Consider this simple statement about motion:

- "If my velocity is zero, I am standing still".

If $X=F(t)$ represent the position, we can restate this mathematically. The theorem is identical to the one in Volume 2 (Chapter 2DC-3), but the constant is a *vector* this time!

Theorem 2.10.5: Constant For Difference Quotients

If a parametric curve defined at the nodes of a partition of interval $[a,b]$ has a zero difference quotient for all nodes in the partition, then this function is constant over the nodes of $[a,b]$; i.e.,

$$\boxed{\frac{\Delta F}{\Delta t}(c)=0\ \Longrightarrow\ F=\ \text{constant}}$$

Proof.

$$\frac{\Delta F}{\Delta x}(c_i)=0\ \Longrightarrow\ F(t_i)-F(t_{i-1})=0\ \Longrightarrow\ F(t_i)=F(t_{i-1})\,.$$

Theorem 2.10.6: Constant For Derivatives

If a differentiable on open interval I function has a zero derivative for all t in I, then this function is constant on I; i.e.,

$$\boxed{F' = 0 \implies F = \text{ constant}}$$

Proof.

To prove that f is constant, it suffices to show that

$$F(a) = F(b)\,,$$

for all a, b in I. Assume $a < b$ and use the *Mean Value Theorem* with interval (a, b) inside the interval I:

$$\frac{F(b) - F(a)}{b - a} = F'(c)\,,$$

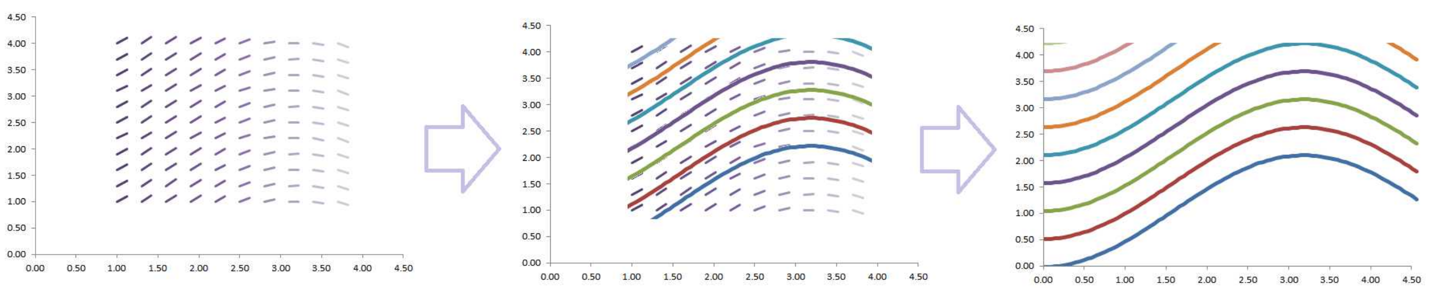
for some c (a, b). This is 0 by assumption. Therefore, we have for all pairs a, b in I:

$$\frac{F(b) - F(a)}{b - a} = 0 \implies F(b) - F(a) = 0 \implies F(a) = F(b)\,.$$

Exercise 2.10.7

What if $F' = 0$ on a union of two intervals?

The problem then becomes one of recovering the function F from it derivative F', the process we have called *anti-differentiation*. In other words, we reconstruct the function from a "field of vectors" (i.e., a vector field):



Now, even if we can recover the function F from it derivative F', there many others with the same derivative, such as $G = F + C$ for any constant real number C. Are there others? No.

Theorem 2.10.8: Anti-differentiation

If two parametric curve defined at the nodes of a partition of interval $[a, b]$ have the same difference quotient, they differ by a constant vector; i.e.,

$$\boxed{\frac{\Delta F}{\Delta x}(c) = \frac{\Delta G}{\Delta x}(c) \implies F(x) - G(x) = \text{ constant}}$$

Theorem 2.10.9: Anti-differentiation

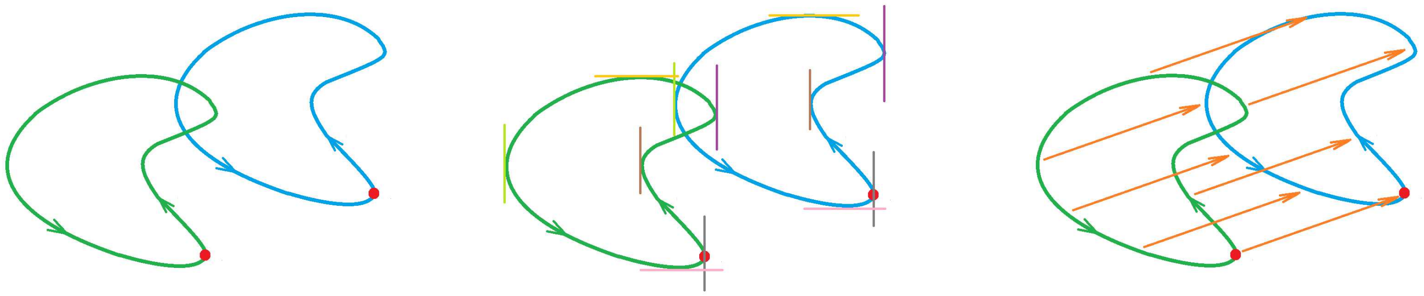
If two differentiable on open interval I parametric curves have the same derivative, they differ by a constant vector; i.e.,

$$\boxed{F'(x) = G'(x) \implies F(x) - G(x) = \text{ constant}}$$

Proof.

Define
$$H(t) = F(t) - G(t)\,.$$
Then, by SR, we have:
$$H'(t) = (F(t) - G(t))' = F'(t) - G'(t) = 0\,,$$
for all x. Then H is constant, by the *Constant Theorem.*

Geometrically, this means that shifting the graph of F gives us the graph of G.



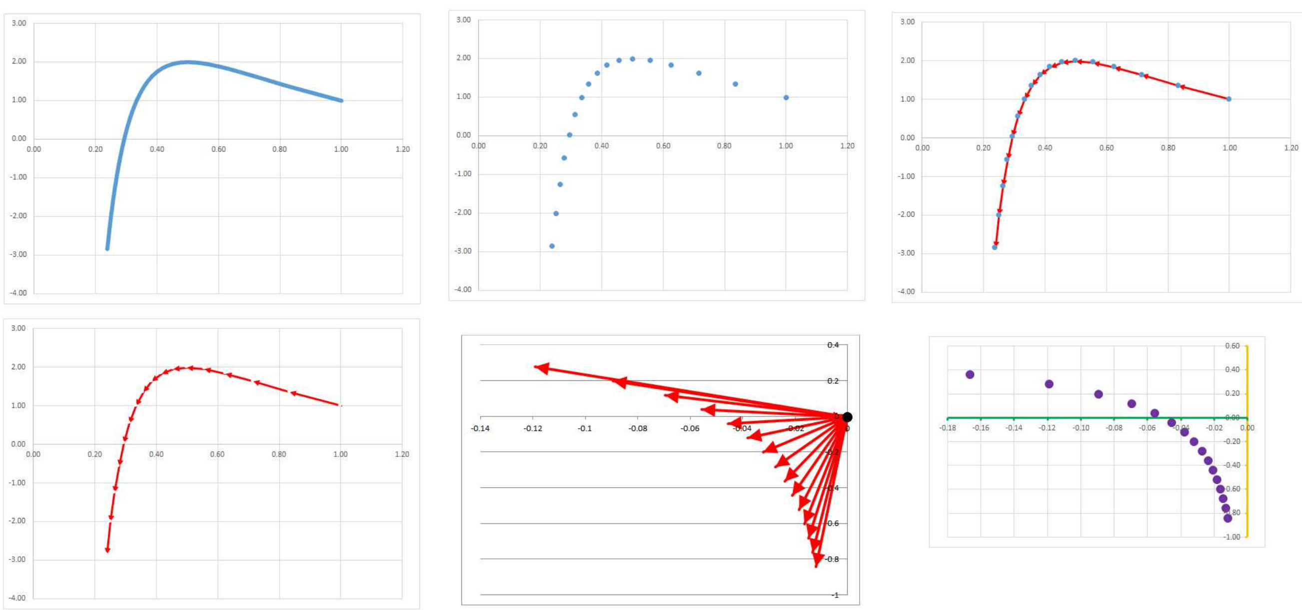
The theorem represents a less obvious fact about motion:

- "If two runners run with the same velocity, their relative location isn't changing".

2.11. Sums and integrals

Let's visualize what it means to differentiate and integrate a parametric curve.

Suppose our parametric curve $X = F(t)$, $a \le t \le b$, is defined at these points: $X_i = F(t_i)$, where $t_{i+1} = t_i + h$, $i = 1, 2, ..., n$. Between these locations we draw the displacement vectors, $D_i = X_iX_{i+1}$:



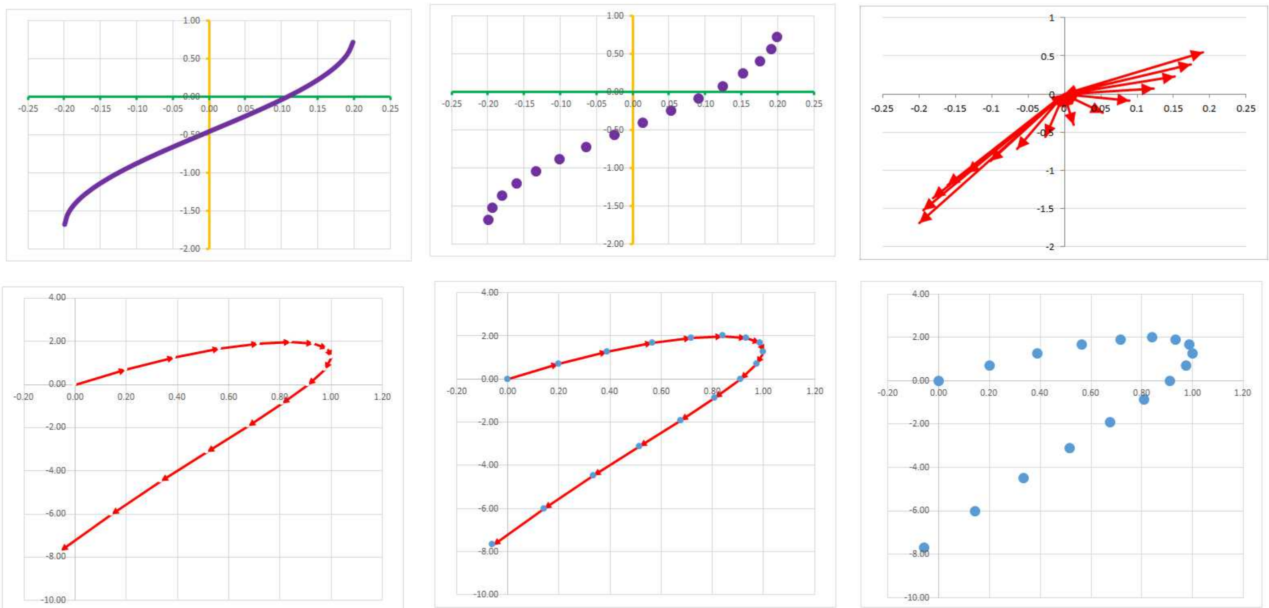
We then move them so that the starting points are all the same, at the origin. We have created the vectors of displacement. By tracing the ends of these vectors we acquire a sequence of "locations" in this new space. The result is the *difference* of $X = F(t)$.

We can also re-scale these vectors all at once: $V_i = \frac{1}{h} D_i$, $i = 1, 2, ..., n$, producing the vectors of average velocity. The result is the *difference quotient* of $X = F(t)$.

If the points came from a *sampled* parametric curve, the limit $h \to 0$ finishes the job: We have a new parametric curve, $Q = G(t)$, the *derivative* of $X = F(t)$.

Now in reverse.

Suppose a parametric curve $Q = G(t),\ a \le t \le b$, is defined at these points: $Q_i = G(t_i)$, where $t_{i+1} = t_i + h,\ i = 1, 2, ..., n$. We represent these “locations” as vectors $D_i = OQ_i$ starting at the same point, the origin:



We think of them as displacements. We arrange them head-to-tail starting at some location X_0 producing a sequence of locations: $X_{i+1} = X_i + D_i,\ i = 1, 2, ..., n$. The result is the *sum* of $Q = G(t)$.

If we instead think of the original vectors as velocities, we re-scale these vectors all at once first: $D_i = L_i h$, producing the displacement vectors. We arrange them head-to-tail starting at some location X_0 producing a sequence of locations: $X_{i+1} = X_i + D_i,\ i = 1, 2, ..., n$. The result is the *Riemann sum* of $Q = G(t)$.

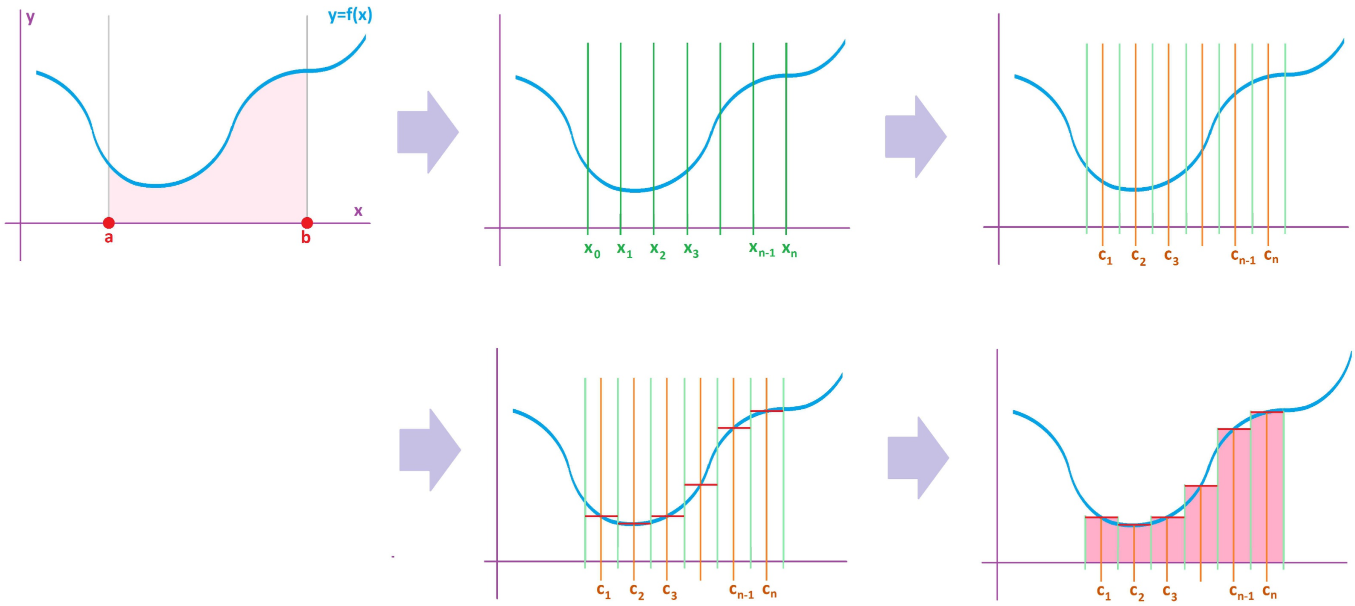
If the points came from a *sampled* parametric curve, the limit $h \to 0$ finishes the job: we have a new parametric curve $X = F(t)$, the *integral* of $Q = G(t)$.

The following is the general setup for sums and Riemann sums.

Suppose we have a parametric curve $X = F(t)$, maybe continuous maybe not, defined on $[a, b],\ a < b$. We start with an *augmented partition* of the interval. Given an integer $n \ge 1$, we have a *partition* of $[a, b]$ into n intervals of possibly different lengths:

$$[t_0, t_1],\ [t_1, t_2],\ ...,\ [t_{n-1}, t_n]\,,$$

with $t_0 = a,\ t_n = b$. The intervals can be possibly reversed!



The points,

$$t_0,\ t_1,\ t_2,\ ...,\ t_{n-1},\ t_n\,,$$

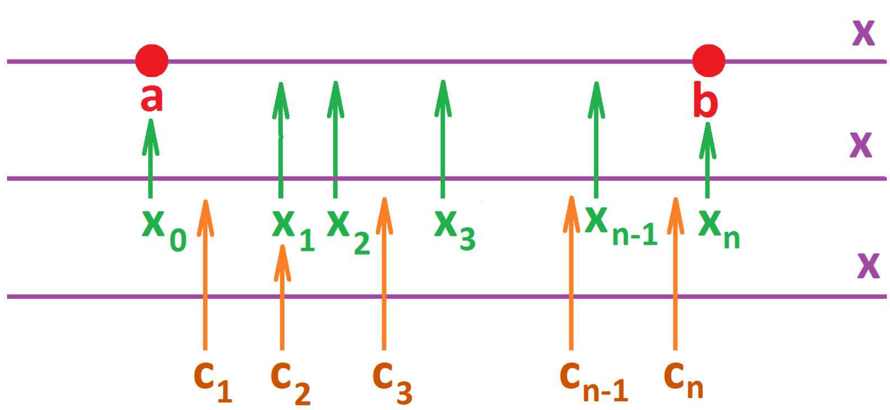
will be called the *nodes* of the partition. The lengths of the intervals are:

$$\Delta t_i = t_i - t_{i-1},\ i = 1, 2, ..., n\,.$$

We are also given the *secondary nodes*:

$$c_1 \text{ in } [t_0, t_1],\ c_2 \text{ in } [t_1, t_2],\ ...,\ c_n \text{ in } [t_{n-1}, t_n]\,.$$

The secondary nodes can be chosen to be the left- or the right-end points of the intervals, or the mid-points, etc.



Definition 2.11.1: sum

The *sum* of a parametric curve $Q = G(t)$ over an augmented partition of an interval $[a, b]$ is defined to be

$$\boxed{\sum_a^b G = G(c_1) + G(c_2) + ... + G(c_n) = \sum_{i=1}^n G(c_i)}$$

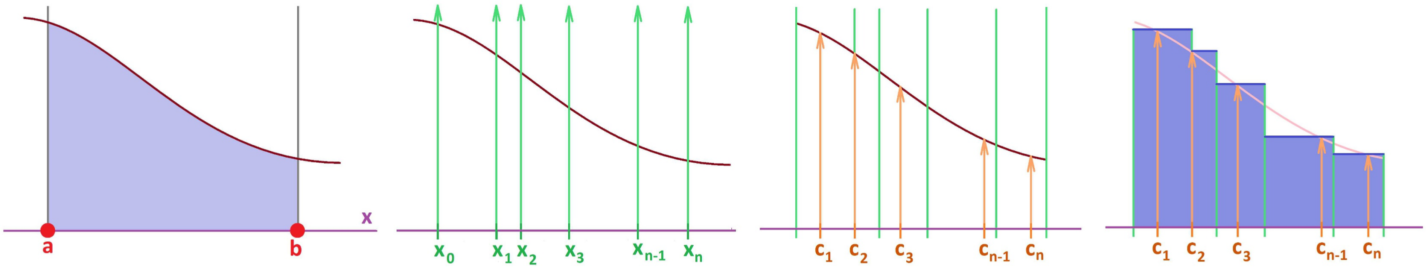
Definition 2.11.2: Riemann sum

The *Riemann sum* of a parametric curve $Q = F(t)$ over an augmented partition

of an interval $[a,b]$ is defined to be

$$\sum_a^b F\,\Delta t = F(c_1)\,\Delta t_1 + F(c_2)\,\Delta t_2 + \dots + F(c_n)\,\Delta t_n = \sum_{i=1}^n F(c_i)\,\Delta t_i$$

The sum may be illustrated as below for the 1-dimensional case:



In higher dimensions, the Riemann sum is a combination of the areas of the rectangles with bases $[t_i,t_{i+1}]$ and heights produced by the *components* of $F(c_i)$.

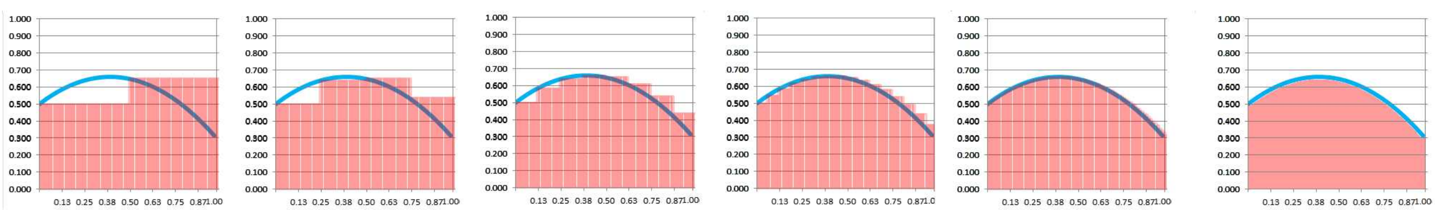
Theorem 2.11.3: Constant Function Rule

Suppose $Q=F(t)$ is constant on $[a,b]$; i.e., $F(t)=C$ for all t in $[a,b]$ and some vector C. Then

$$\sum_a^b F\,\Delta t = C(b-a)$$

Now, in order improve these approximations, we *refine* the partition and keep refining; i.e., we have simultaneously:

$$n\to\infty \text{ and } \Delta t_i \to 0\,.$$



Specifically, we define the *mesh of a partition* P as follows:

$$|P| = \max_i \Delta t_i\,.$$

It is a measure of "refinement" of P.

Definition 2.11.4: Riemann integral

The *Riemann integral* of a parametric curve F over interval $[a,b]$ is defined to be the limit of a sequence of its Riemann sums over a sequence partitions P_k with their mesh approaching 0 as $k\to\infty$. When all these limits exist and are all equal to each other, F is called *integrable* over $[a,b]$ and the limit is denoted as follows:

$$\int_a^b F\,dt = \lim_{k\to\infty}\sum_a^b F_k\,\Delta t$$

where F_k is F sampled at the secondary nodes of P_k. It is also called the *definite integral* of F. The interval $[a, b]$ is the *domain of integration*.

Theorem 2.11.5: Constant Integral Rule

Suppose F is constant on $[a, b]$, i.e., $F(t) = C$ for all t in $[a, b]$ and some vector C. Then F is integrable on $[a, b]$ and

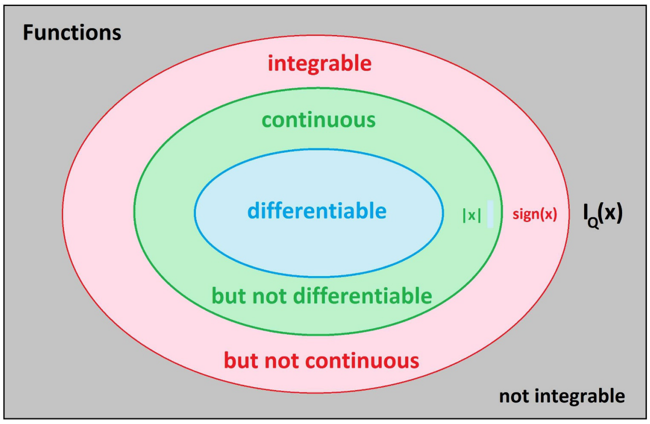
$$\boxed{\int_a^b F\, dt = C(b - a)}$$

The following result proves that our definition makes sense for a large class of parametric curves.

Theorem 2.11.6: Cont. => Integr.

All continuous parametric curves on $[a, b]$ are integrable on $[a, b]$.

Examples can be taken from the numerical case (Volumes 2 and 3):



Theorem 2.11.7: Negative Integral

The Riemann integral of a parametric curve F over interval $[b, a]$ is equal to the negative of the integral over $[a, b]$:

$$\boxed{\int_b^a F\, dt = -\int_a^b F\, dt}$$

Exercise 2.11.8

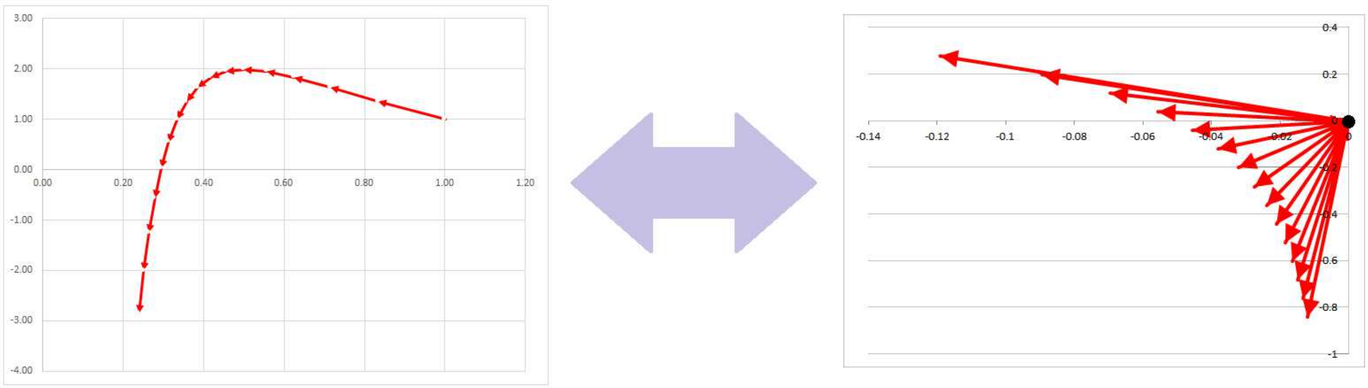
Prove the theorem.

2.12. The Fundamental Theorem of Calculus

What is the opposite of subtraction? Addition. Then the opposite of the *difference* is the *sum*:

$$\left.\begin{array}{ll}\text{difference, } \Delta F: & \text{vector subtraction} \\ \\ \text{sum, } \displaystyle\sum_a^t F: & \text{vector addition}\end{array}\right\}\text{opposite!}$$

The two operations *cancel* each other!

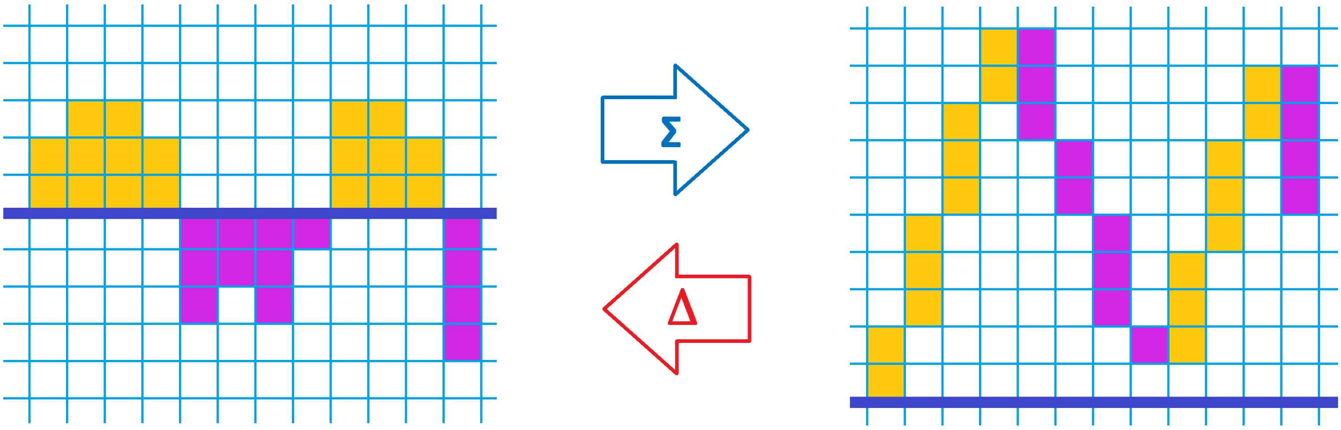


Theorem 2.12.1: Fundamental Theorem of Discrete Calculus of Degree 1

Suppose F is a parametric curve defined at the nodes of a partition of an interval $[a,b]$ and suppose a is a node of the partition. Then, for each node t, we have:

$$\boxed{\sum_a^t (\Delta F) = F(t) - F(a)}$$

The components of the differences and the sum are the differences and sum of component functions as defined in Volume 3 (Chapter 3IC-4). The relation we observed then now appears separately for each component:

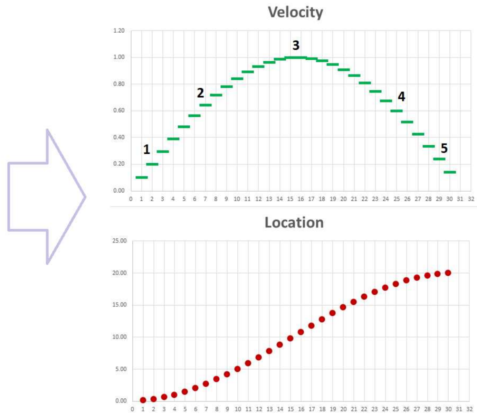


We take this idea further and examine the *fundamental* relation between the Riemann sums and the difference quotients. It is only slightly more complicated.

In our vector-valued setting, let's take another look at the computations of motion.

First, this is how we use the velocity function to acquire the displacement. Then we discover that each of the values of the latter is a Riemann sum of the former!

time min	velocity miles/min	location miles
0		0.00
1	0.10	0.10
2	0.20	0.30
3	0.30	0.59
4	0.39	0.98
5	0.48	1.46
6	0.56	2.03
7	0.64	2.67
8	0.72	3.39
9	0.78	4.17
10	0.84	5.01
11	0.89	5.91
12	0.93	6.84
13	0.96	7.80
14	0.99	8.79
15	1.00	9.78
16	1.00	10.78
17	0.99	11.77
18	0.97	12.75
19	0.95	13.70
20	0.91	14.60
21	0.86	15.47
22	0.81	16.28
23	0.75	17.02
24	0.68	17.70
25	0.60	18.30
26	0.52	18.81
27	0.43	19.24
28	0.33	19.57
29	0.24	19.81
30	0.14	19.95



Indeed, suppose we have a parametric curve G defined at the secondary nodes of a partition of an interval $[a,b]$, i.e., a 1-form with values in $\mathbf{R}^n$. Its Riemann sum defines a new parametric curve Γ on the nodes, i.e., a 0-form. It can be computed directly:

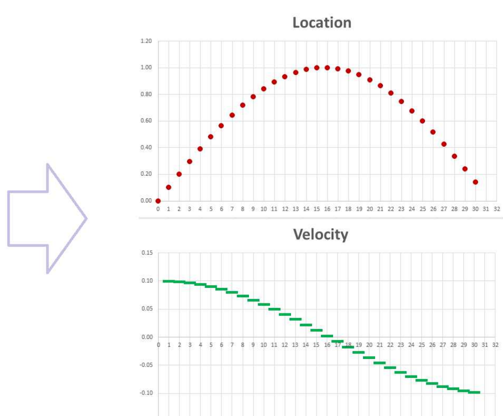
$$\Gamma(t_k) = \sum_a^{t_k} G\,\Delta t\,,$$

or recursively:

$$\Gamma(t_{k+1}) = \Gamma(t_k) + G(c_k)\,\Delta t_k\,.$$

Second, this is how we use the parametric curve of location to acquire the velocity. Then we discover that each of the values of the latter is a difference quotient of the former!

time min	location miles	velocity miles/min
0	0.00	
1	0.10	0.10
2	0.20	0.10
3	0.30	0.10
4	0.39	0.09
5	0.48	0.09
6	0.56	0.09
7	0.64	0.08
8	0.72	0.07
9	0.78	0.07
10	0.84	0.06
11	0.89	0.05
12	0.93	0.04
13	0.96	0.03
14	0.99	0.02
15	1.00	0.01
16	1.00	0.00
17	0.99	-0.01
18	0.97	-0.02
19	0.95	-0.03
20	0.91	-0.04
21	0.86	-0.05
22	0.81	-0.05
23	0.75	-0.06
24	0.68	-0.07
25	0.60	-0.08
26	0.52	-0.08
27	0.43	-0.09
28	0.33	-0.09
29	0.24	-0.10
30	0.14	-0.10



Indeed, suppose we have a parametric curve Φ at the primary nodes of a partition of an interval $[a,b]$, i.e., a 0-form with values in $\mathbf{R}^n$. Its difference quotient is computed over each segment of the partition $[t_{k-1},t_k]$, $k=1,2,...,n$, and assigned to the corresponding secondary node c_k. This defines a new parametric curve F on the secondary nodes, i.e., a 1-form. It is computed by the familiar formula:

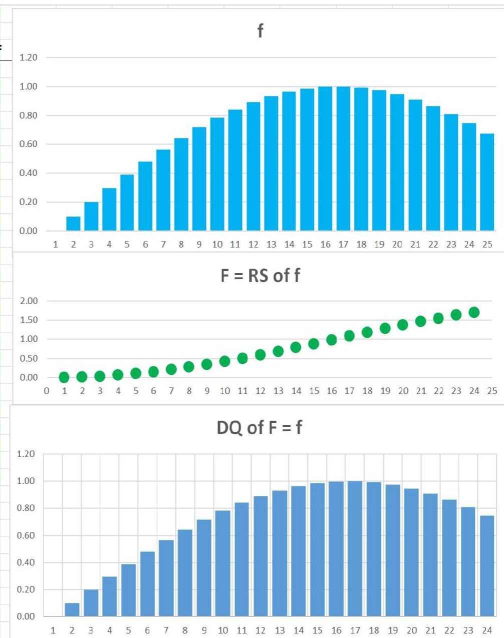
$$F(c_k) = \frac{\Phi(t_k)-\Phi(t_{k-1})}{\Delta t_k}\,.$$

We realize that the difference quotients of the Riemann sums give us the original parametric curve; we just substitute the recursive formula for the latter into the last formula for the former: $\Phi=\Gamma$. Then we have:

$$\begin{aligned} F(c_k) &= \frac{\Gamma(t_k)-\Gamma(t_{k-1})}{\Delta t_k} \\ &= \frac{G(c_k)\Delta t_k}{\Delta t_k} \\ &= G(c_k)\,. \end{aligned}$$

The result is seen in the spreadsheet below:

h			
0.10			
x	f	F = RS of f	DQ of F = f
0.00	0.00	0.00	
0.10	0.10	0.00	0.00
0.20	0.20	0.01	0.10
0.30	0.30	0.03	0.20
0.40	0.39	0.06	0.30
0.50	0.48	0.10	0.39
0.60	0.56	0.15	0.48
0.70	0.64	0.20	0.56
0.80	0.72	0.27	0.64
0.90	0.78	0.34	0.72
1.00	0.84	0.42	0.78
1.10	0.89	0.50	0.84
1.20	0.93	0.59	0.89
1.30	0.96	0.68	0.93
1.40	0.99	0.78	0.96
1.50	1.00	0.88	0.99
1.60	1.00	0.98	1.00
1.70	0.99	1.08	1.00
1.80	0.97	1.18	0.99
1.90	0.95	1.27	0.97
2.00	0.91	1.37	0.95
2.10	0.86	1.46	0.91
2.20	0.81	1.55	0.86
2.30	0.75	1.63	0.81
2.40	0.68	1.70	0.75



Now, vice versa, the Riemann sums of the difference quotients give us the original – up to a constant vector – parametric curve; we just substitute the last formula for the former into the recursive formula for the latter: $G(c_k) = F(c_k)$. If we assume that $\Gamma(t_k) = \Phi(t_k) + C$ for all k, the we conclude:

$$\begin{aligned}
\Gamma(t_k) &= \Gamma(t_{k-1}) + F(c_k)\Delta t_k \\
&= \Gamma(t_{k-1}) + \frac{\Phi(t_k) - \Phi(t_{k-1})}{\Delta t_k}\Delta t_k \\
&= \Gamma(t_{k-1}) + \Phi(t_k) - F(t_{k-1}) \\
&= \Phi(t_k) + C\,.
\end{aligned}$$

Let's summarize what we have proven – for discretely defined parametric curves, i.e., a 0- and 1-forms with values in $\mathbf{R}^n$.

Definition 2.12.2: difference quotient

Suppose Φ is defined at the primary nodes $t_k,\ k = 0, 1, 2, ..., n$, of the partition. Then the *difference quotient* F of Φ is defined at the secondary nodes of the partition by:

$$\boxed{F(c_k) = \frac{\Phi(t_k) - \Phi(t_{k-1})}{\Delta t_k}}$$

It is denoted as follows

$$\boxed{F = \frac{\Delta \Phi}{\Delta t}}$$

Definition 2.12.3: Riemann sum

Suppose G is defined at the secondary nodes $c_k,\ k = 1, 2, ..., n$, of the partition. Then the *Riemann sum* of G is defined recursively at the primary nodes of the partition by:

$$\boxed{\Gamma(t_k) = \Gamma(t_{k-1}) + G(c_k)\Delta t_k}$$

It is denoted as follows:

$$\boxed{\Gamma = \sum_a^t G\,\Delta t}$$

Theorem 2.12.4: Fundamental Theorem of Discrete Calculus

(1) The difference quotient of the Riemann sum of G is G:

$$\boxed{\frac{\Delta\left(\sum_a^t G\,\Delta t\right)}{\Delta t} = G}$$

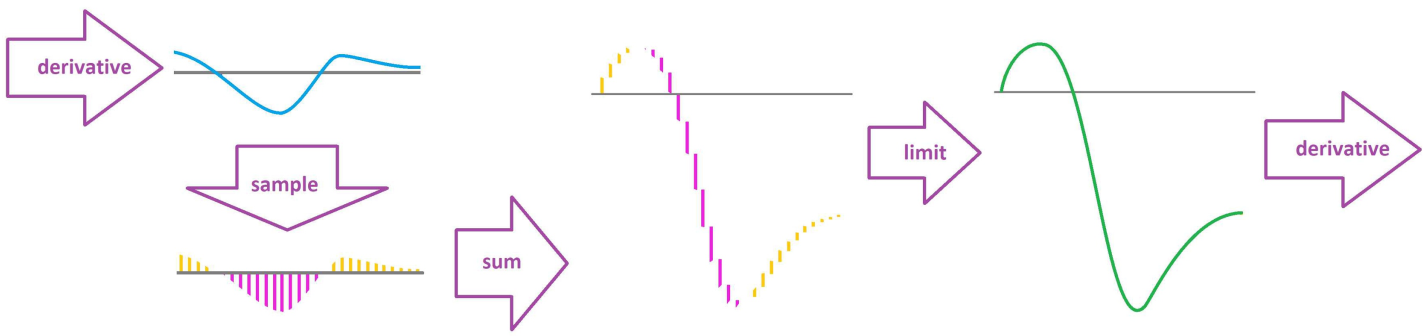
(2) The Riemann sum of the difference quotient of Φ is $\Phi + C$, where C is a constant:

$$\boxed{\sum_a^t \left(\frac{\Delta\Phi}{\Delta t}\right)\Delta t = \Phi(t) + C}$$

The two operations *cancel* each other! The result shouldn't be surprising considering the operations involved:

$$\left.\begin{array}{llcl}\text{difference quotient, } \dfrac{\Delta F}{\Delta t}: & \text{vector subtraction} & \to & \text{scalar division}\\[3ex] \text{Riemann sum, } \displaystyle\sum_a^t F\,\Delta t: & \text{vector addition} & \leftarrow & \text{scalar multiplication}\end{array}\right\}\text{opposite!}$$

Now the *continuous case*. We just take the above relations and let Δt approach 0. We "zoom out" on the graph:



Definition 2.12.5: antiderivative

A parametric curve $X = F(t)$ is called an *antiderivative* of a parametric curve $Q = G(t)$ if $F' = G$. It is denoted as follows:

$$\boxed{G = \int F\,dt}$$

We use "an" because there are many antiderivatives for each parametric curve.

Theorem 2.12.6: Fundamental Theorem of Calculus

1. *Given a continuous parametric curve $Q = F(t)$ on $[a, b]$, the function defined by*

$$\boxed{\Phi(x) = \int_a^x F\,dt}$$

is an antiderivative of F on (a, b).

2. *For a continuous parametric curve $Q = F(t)$ on $[a, b]$ and any of its antiderivatives Φ, we have*

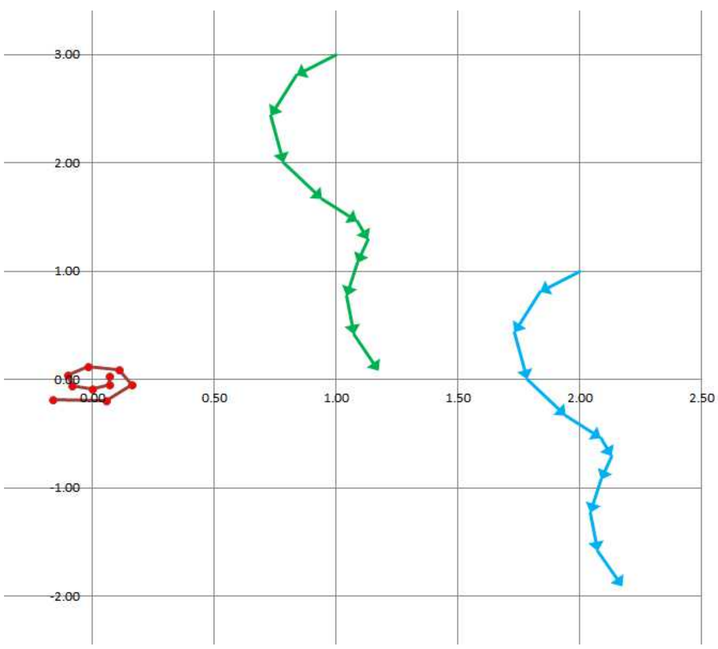
$$\boxed{\int_a^b F\,dt = \Phi(b) - \Phi(a)}$$

According to the Fundamental Theorem, the operations of differentiation and integration cancel each other:

$$\begin{array}{ccccccccc} F & \to & \boxed{\int (\quad)\,dt} & \to & \Phi & \to & \boxed{\frac{d}{dt}(\quad)} & \to & F. \\ \Phi & \to & \boxed{\frac{d}{dt}(\quad)} & \to & F & \to & \boxed{\int (\quad)\,dt} & \to & \Phi + C. \end{array}$$

Corollary 2.12.7: Antiderivative Plus Constant

If G is an antiderivative of F then so is $G + C$, where C is any constant vector.

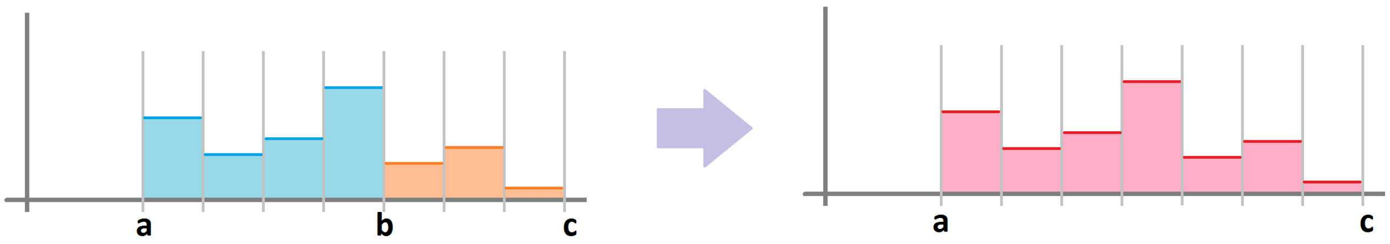


Thus, of $X = F(t)$ and $X = G(t)$ are two antiderivatives of the same parametric curve, then not only their paths are the same, shifted by a fixed vector, but also the two locations $F(t)$ and $G(t)$ at the exact moment of time are also separated by that vector.

2.13. Algebraic properties of sums and integrals

Just as in dimension 1 presented in Volume 1 (Chapter 1PC-1), algebraic operations on functions produce algebraic operations on their sums. The properties for integrals follow from the corresponding properties of the sums and the rules of limits.

If we proceed to the adjacent interval, we can just continue to add terms of the sum (one component shown):



Theorem 2.13.1: Additivity Rule

Suppose F is a vector-valued function (parametric curve) defined at the nodes of a partition. Then, for any nodes a, b, c, we have:

$$\boxed{\sum_a^b F + \sum_b^c F = \sum_a^c F}$$

over partitions of $[a, b]$, $[b, c]$, and $[a, c]$ respectively.

Theorem 2.13.2: Additivity Rule

Suppose F is a vector-valued function (parametric curve) defined at the nodes of a partition. Then, for any nodes a, b, c, we have:

$$\boxed{\sum_a^b F\,\Delta t + \sum_b^c F\,\Delta t = \sum_a^c F\,\Delta t}$$

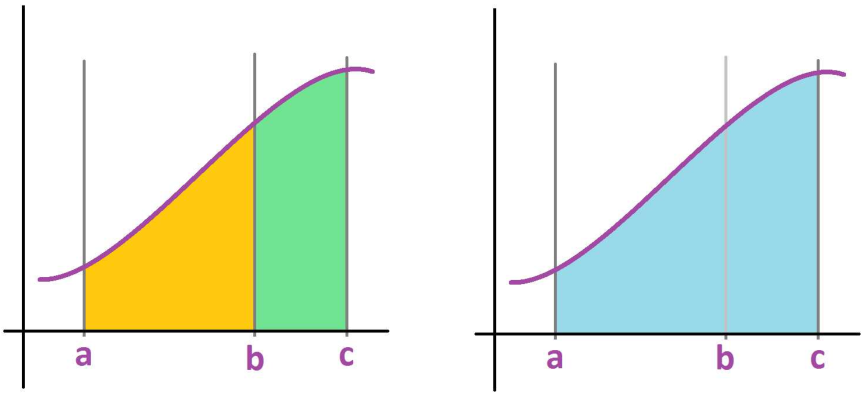
over partitions of $[a, b]$, $[b, c]$, and $[a, c]$ respectively.

Theorem 2.13.3: Additivity Rule

Suppose a vector-valued function (parametric curve) F is integrable over $[a, b]$ and over $[b, c]$. Then F is integrable $[a, c]$ and we have:

$$\boxed{\int_a^b F\,dt + \int_b^c F\,dt = \int_a^c F\,dt}$$

The area interpretation of additivity of integrals is the same:



With the motion interpretation, we have:

$$\begin{aligned} & \text{displacement during the 1st hour} \\ + & \text{ displacement during the 2nd hour} \\ = & \text{ displacement the two hours.} \end{aligned}$$

Theorem 2.13.4: Estimate Rule

Suppose F is a vector-valued function (parametric curve) defined at the nodes of a partition. Then, for any nodes a, b with $a < b$, if

$$||F(x)|| \leq M\,,$$

for all t with $a \leq t \leq b$, then

$$\left|\left|\sum_a^b F\,\Delta t\right|\right| \leq M(b-a)\,.$$

Theorem 2.13.5: Estimate Rule

Suppose F is an integrable vector-valued function (parametric curve) over $[a, b]$. Then, if $a < b$ and

$$||F(x)|| \leq M\,,$$

for all t with $a \leq t \leq b$, we have:

$$\left|\left|\int_a^b F\,dt\right|\right| \leq M(b-a)\,.$$

Note that the statement of the theorem still holds even if F is integrable over $[b, a]$ with $b < a$, etc. This implies the following important result.

Theorem 2.13.6: Local Integrability

If a vector-valued function (parametric curve) F is integrable over $[a, b]$ then it is also integrable over any $[a', b']$ with $a \leq a' < b' \leq b$.

The following is another important corollary.

Theorem 2.13.7: Cont. => Integr.

All piecewise continuous parametric curves are integrable.

These are the algebraic properties.

Theorem 2.13.8: Constant Multiple Rule For Sums

Suppose F is a vector-valued function (parametric curve) defined at the nodes of a partition. Then, for any nodes a, b and any real c, we have:

$$\boxed{\sum_a^b (c \cdot F) = c \cdot \sum_a^b F}$$

Theorem 2.13.9: Constant Multiple Rule For Riemann Sums

Suppose F is a vector-valued function (parametric curve) defined at the nodes of a partition. Then, for any nodes a, b and any real c, we have:

$$\boxed{\sum_a^b (c \cdot F)\,\Delta t = c \cdot \sum_a^b F\,\Delta t}$$

Theorem 2.13.10: Constant Multiple Rule For Integrals

Suppose a vector-valued function (parametric curve) F is integrable over $[a, b]$. Then so is $c \cdot f$ for any real c and we have:

$$\boxed{\int_a^b (c \cdot F)\,dt = c \cdot \int_a^b F\,dt}$$

Theorem 2.13.11: Sum Rule For Sums

Suppose F and G are vector-valued functions (parametric curves) defined at the nodes of a partition. Then, for any nodes a, b, we have:

$$\boxed{\sum_a^b (F + G) = \sum_a^b F + \sum_a^b G}$$

Theorem 2.13.12: Sum Rule For Riemann Sums

Suppose F and G are vector-valued functions (parametric curves) defined at the nodes of a partition. Then, for any nodes a, b, we have:

$$\boxed{\sum_a^b (F + G)\ \Delta t = \sum_a^b F\,\Delta t + \sum_a^b G\,\Delta t}$$

Theorem 2.13.13: Sum Rule For Integrals

Suppose vector-valued functions (parametric curves) F and G are integrable over

$[a,b]$. *Then so is* $F+G$ *and we have:*

$$\boxed{\int_a^b (F+G)\,dt = \int_a^b F\,dt + \int_a^b G\,dt}$$

2.14. The rate of change of the rate of change

As we know, this is about the second difference quotient and the second derivative.

While the original parametric curve was defined at the nodes of a partition, its difference and the difference quotient are defined at the secondary nodes. Following the ideas from Volume 2 (Chapter 2DC-4), we treat the latter as functions that also have their own difference and difference quotient. What is the partition then? We saw earlier in the chapter how this idea is implemented in order to derive the acceleration from the velocity.

First, also following Chapter 2DC-4, we notice that the difference of the difference is simply the difference that skips a node. That's why the second difference doesn't provide us – in this 1-dimensional setting – with any meaningful information. We will limit our attention to the *difference quotients*.

We need *three* values of the function F at the nodes. We then compute the difference quotients along the two intervals and place the results within the corresponding edge. Finally, the same operation is carried out for these two values:

$$\begin{array}{ccccccc}
- & F(t_1) & --- & F(t_2) & --- & F(t_3) & - \\
- & -\bullet- & \dfrac{\Delta F}{\Delta t_2} & -\bullet- & \dfrac{\Delta F}{\Delta t_3} & -\bullet- & - \\
- & -\bullet- & --- & \dfrac{\frac{\Delta F}{\Delta t_3}-\frac{\Delta F}{\Delta t_2}}{c_3-c_2} & --- & -\bullet- & - \\
 & t_1 & c_2 & t_2 & c_3 & t_3 &
\end{array}$$

This is all *vector algebra*!

The function defined above represents the directions of the lines connecting the values of F at consecutive nodes of the partition. In particular, when the location is represented by a function known only at the nodes of the partition, the (average) velocity is then found in this manner. It is now especially important that we have utilized the secondary nodes as the inputs of the new function. Indeed, we can now carry out a similar construction with this function and find the (average) *acceleration*!

We have now a new *augmented partition*, of what? The interval is

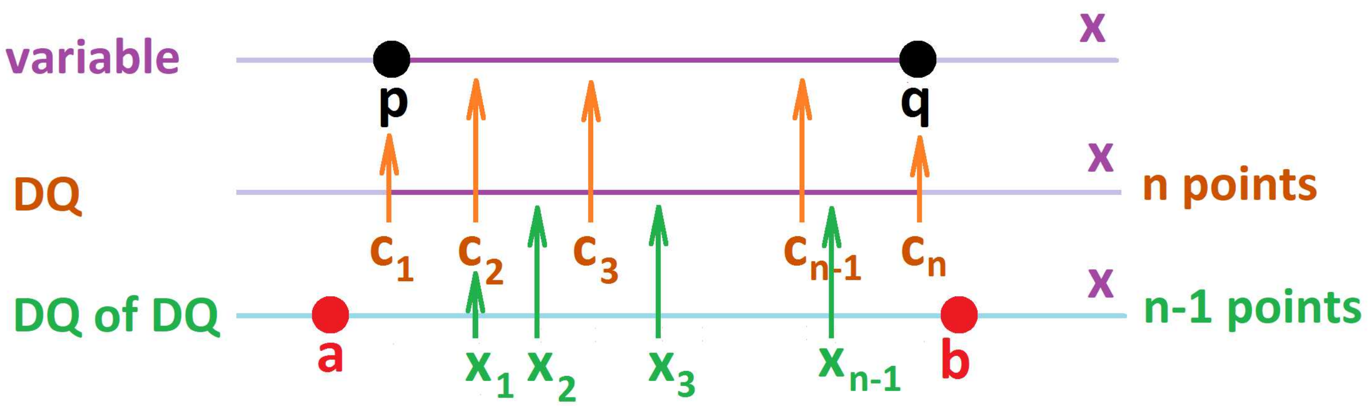
$$[p,q], \quad \text{with } p=c_0 \ \text{ and } \ q=c_n\,.$$

We partition it into $n-1$ intervals with the help of the nodes that used to be the secondary nodes in the last partition:

$$p=c_1,\ c_2,\ c_3,\ \ldots,\ c_{n-1},\ c_n=b\,.$$

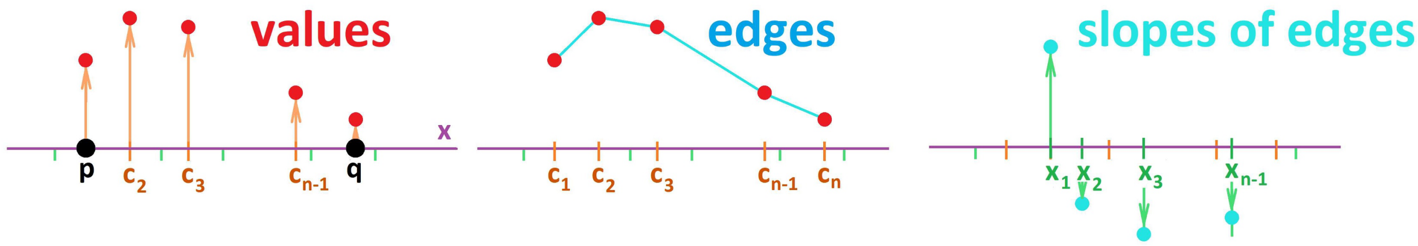
When $c_k=c_{k+1}$, we exclude this "interval" from the new partition. Now, what are the secondary nodes? The primary nodes of the last partition of course! Indeed, we have:

$$t_1 \text{ in } [c_1,c_2],\ t_2 \text{ in } [c_2,c_3],\ \ldots,\ t_{n-1} \text{ in } [c_{n-1},c_n]\,.$$



We apply the same construction to this partition to the function $G = \frac{\Delta F}{\Delta t}$. The difference quotient function of F is defined at the secondary nodes of the new partition (the primary nodes of the old partition) by the same formula:

$$\frac{\Delta G}{\Delta t}(t_k) = \frac{G(c_{k+1}) - G(c_k)}{c_{k+1} - c_k}.$$



Definition 2.14.1: second difference quotient

The *second difference quotient* of $X = F(t)$ is defined at the primary nodes of the partition by:

$$\boxed{\frac{\Delta^2 F}{\Delta t^2}(t_k) = \frac{\frac{\Delta F}{\Delta t}(c_{k+1}) - \frac{\Delta F}{\Delta t}(c_k)}{c_{k+1} - c_k},\quad k = 1, 2, ..., n-1}$$

It is then a vector-valued 0-form.

Example 2.14.2: circle

Let's find the second difference quotient of the standard parametrization of the circle,

$$X(t) = <\cos t,\ \sin t>\ .$$

It was considered previously in this chapter:

$$\frac{\Delta X}{\Delta t}(c) = \frac{\sin(h/2)}{h/2} < -\sin c,\ \cos c>\ .$$

Here we have a mid-point partition for the interval, say, $[-\pi/2, \pi/2]$, in the t-axis:

- The nodes are $x = a, a + h, ...$ and
- The secondary nodes are $c = a + h/2, ...$.

Over the same mid-point partition, we have the second difference quotient:

$$\frac{\Delta^2 X}{\Delta t^2}(a) = -\left(\frac{\sin(h/2)}{h/2}\right)^2 < \cos a,\ \sin a>\ .$$

This vector points toward the center of the circle!

The construction of the difference quotient is repeatedly used for approximations and simulations which is followed, when necessary, by taking its *limit*. The result is the derivative and the second derivative.

The derivative of a differentiable parametric curve is also a parametric curve and can also be differentiated.

The notation is identical to that for numerical functions:

Repeated derivatives

function	F	$F^{(0)}$	
first derivative	F'	$F^{(1)}$	$\frac{dF}{dt}$
second derivative	$F''=(F')'$	$F^{(2)}=\left(F^{(1)}\right)'$	$\frac{d^2F}{dt^2}=\frac{d}{dt}\left(\frac{dF}{dt}\right)$
third derivative	$F'''=(F'')'$	$F^{(3)}=\left(F^{(2)}\right)'$	$\frac{d^3F}{dt^3}=\frac{d}{dt}\left(\frac{d^2F}{dt^2}\right)$
...		...	...
nth derivative		$F^{(n)}=\left(F^{(n-1)}\right)'$	$\frac{d^nF}{dt^n}=\frac{d}{dt}\left(\frac{d^{n-1}F}{dt^{n-1}}\right)$
...		...	...

Thus, a given differentiable function may produce a *sequence of parametric curves*:

$$F \to \boxed{\frac{d}{dt}} \to F' \to \boxed{\frac{d}{dt}} \to F'' \to \ \dots \ \to F^{(n)} \to \ \dots$$

provided the outcome of each step is differentiable as well. In the abbreviated form the sequence is:

$$F \xrightarrow{\frac{d}{dt}} F' \xrightarrow{\frac{d}{dt}} F'' \xrightarrow{\frac{d}{dt}} \dots \xrightarrow{\frac{d}{dt}} F^{(n)} \xrightarrow{\frac{d}{dt}} \dots$$

Warning!

We will see in Chapter 3 that the function and its derivative are two animals of very different breeds in higher dimensions.

Definition 2.14.3: multiply differentiable function

A parametric curve F is called *twice, thrice, ..., n times differentiable* when the corresponding derivatives,

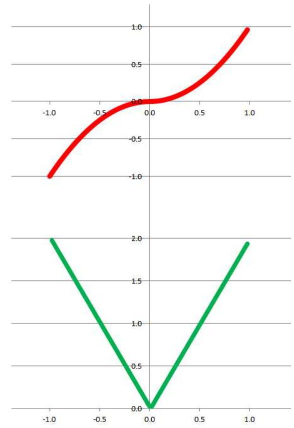
$$F', F'', F''', ..., F^{(n)}\,,$$

exist. When the derivatives exist for all n, we call the function *smooth*.

Example 2.14.4: repeated differentiability

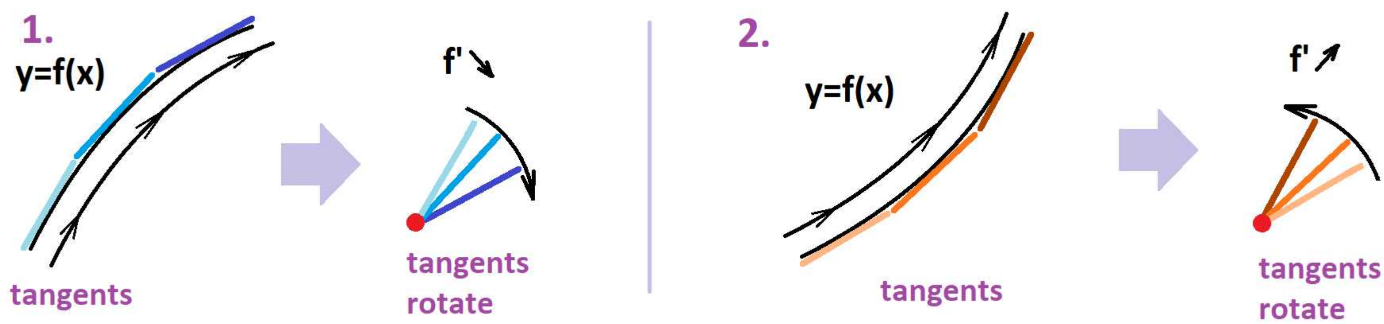
Examples of functions that are differentiable but not twice differentiable come from Volume 2 (Chapter 2DC-3):

$$f(x)=\begin{cases} -x^2 & \text{if } x<0,\\ x^2 & \text{if } x\ge 0.\end{cases}$$



What is the geometric meaning of these higher derivatives?

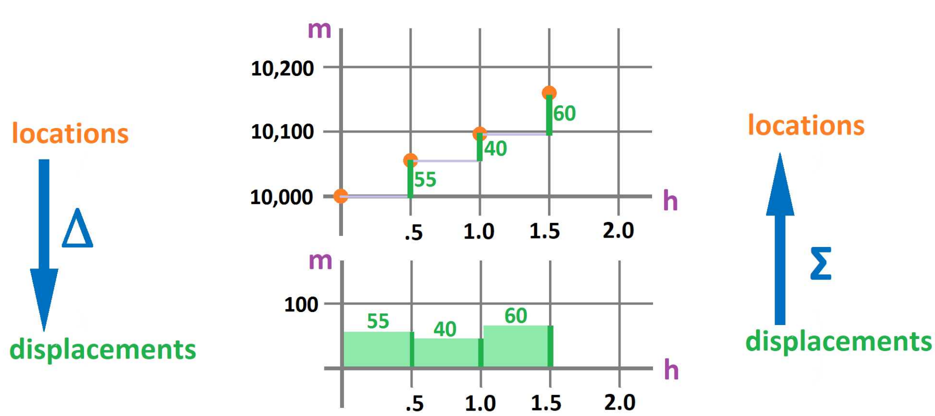
The first derivative represents the direction and the magnitude of change of the *vector-valued* function. Then the second derivative represents the rate of change of these directions and these magnitudes. Notice how changing slopes are seen as rotating tangents:



It other words, this is the *acceleration*.

2.15. Reversing differentiation: antiderivatives

Now in the vector algebra environment, we come back to the same question as in Volume 2 (Chapter 2DC-3): when we know the velocity at every moment of time, how do we find the location? How do we "reverse" the effect of differentiation on a function?



For a parametric curve $X = G(t)$ defined at the secondary nodes, c_k, of a partition, the answer remains the same; we have an almost identical *recursive formula* for the parametric curve $X = F(t)$ defined at the nodes, t_k, of the partition:

$$F(t_{k+1}) = F(t_k + \Delta t_k) = F(t_k) + G(c_k)\Delta t_k\,.$$

Then,

$$\frac{\Delta F}{\Delta t_k} = G(c_k)\,.$$

This is the difference quotient of F is G.

What about the derivative? With the complexity added by the limit, there is no formula, even recursive.

First, let's review how the Sum Rule, the Constant Multiple Rule, and the Chain Rule are represented as diagrams:

$$\begin{array}{ccccccccc}
F,G & \xrightarrow{\frac{d}{dt}} & F',G' & F & \xrightarrow{\frac{d}{dt}} & F' & F,G & \xrightarrow{\frac{d}{dt}} & F',G' \\
\Big\downarrow + & SR & \Big\downarrow + & \Big\downarrow \cdot c & CMR & \Big\downarrow \cdot c & \Big\downarrow \circ & CR & \Big\downarrow \circ \\
F+G & \xrightarrow{\frac{d}{dt}} & (F+G)' = F'+G' & cF & \xrightarrow{\frac{d}{dt}} & (cF)' = cF' & F\circ G & \xrightarrow{\frac{d}{dt}} & (F\circ G)' = F'\circ G'
\end{array}$$

In the first diagram, we start with a pair of functions at the top left and then we proceed in two ways:

- Right: differentiate. Down: add the results.
- Down: add them. Right: differentiate the result.

The result is the same!

Warning!

Neither the Product Rule nor the Quotient Rule has such an interpretation.

Now, anti-differentiation is meant to "reverse" the effect of differentiation on a function just as before. It is similar to the *inverse* of a function.

The importance of this "inverse" problem stems from the need to find location from velocity or velocity from acceleration. For example, this is what we derive from our experience with differentiation of parametric curves:

- The acceleration is constant, $A \Longrightarrow$
- The velocity is a linear function, $V(t) = At + B \Longrightarrow$
- The location is a quadratic function, $X(t) = At^2/2 + Bt + C$.

Here B and C are the initial velocity and the initial position respectively.

We illustrate the idea with a diagram:

$$\begin{array}{ccccc}
< t^2,\ t^3 > & \to & \boxed{\frac{d}{dt}} & \to & < 2t,\ 3t^2 > \\
< 2t,\ 3t^2 > & \to & \boxed{\left(\frac{d}{dt}\right)^{-1}} & \to & < t^2 + C,\ t^3 + K >
\end{array}$$

As a function, $\frac{d}{dt}$ isn't one-to-one!

We will need the *rules of anti-differentiation*.

First, consider SR: $(F+G)' = F' + G'$. Let's read from right to left.

Theorem 2.15.1: Sum Rule

If

- *F is an antiderivative of P and*
- *G is an antiderivative of Q,*

then

- *$F+G$ is an antiderivative of $P+Q$.*

Proof.

We apply SR:

$$(F(x)+G(x))' = F'(x)+G'(x) = P(x)+Q(x)\,.$$

Similarly, we acquire the following.

Theorem 2.15.2: Constant Multiple Rule

If
- *F is and antiderivative of P and*
- *c is a constant,*

then
- *cF is an antiderivative of cP.*

Proof.

We apply CMR:

$$(cF(x))' = cF'(x) = cP(x)\,.$$

Theorem 2.15.3: Linear Chain Rule

If
- *F is and antiderivative of P and*
- *$a \neq 0,\ b$ are constants.*
- *Then $\frac{1}{a}F(ax+b)$ is an antiderivative of $P(ax+b)$.*

Proof.

We apply CMR and CR:

$$\left(\tfrac{1}{a}F(ax+b)\right)' = \tfrac{1}{a}\left(F(ax+b)\right)' = \tfrac{1}{a}aF'(ax+b) = F'(ax+b) = P(ax+b)\,.$$

These diagrams illustrate how differentiation and anti-differentiation undo each other:

$$\begin{array}{ccccccccc}
F & \to & \boxed{\left(\frac{d}{dt}\right)^{-1}} & \to & F & \to & \boxed{\frac{d}{dt}} & \to & F\\
F & \to & \boxed{\frac{d}{dt}} & \to & F & \to & \boxed{\left(\frac{d}{dt}\right)^{-1}} & \to & F+C
\end{array}$$

Just as with the numerical functions, there is a constant of integration but this time it is a *vector*.

We restate the rules.

Sum Rule:

$$\int (F+G)dt = \int F\,dt + \int G\,dt\,.$$

Constant Multiple Rule:

$$\int (cF)dt = c\int F\,dt\,.$$

Linear Chain Rule:

$$\int F(mt+b)dt = \tfrac{1}{m}\int F(u)\,du\Big|_{u=mt+b}\,.$$

At least, we have these two diagrams to illustrate the interaction of antiderivatives with algebra:

$$\begin{tikzcd} P, Q \arrow[d, "+"] & F', G' \arrow[l, "\int"'] \arrow[d, "+"] \\ P+Q & F'+G' \arrow[l, "\int"'] \end{tikzcd} \qquad \begin{tikzcd} P \arrow[d, "\cdot c"] & F' \arrow[l, "\int"'] \arrow[d, "\cdot c"] \\ cP & cF' \arrow[l, "\int"'] \end{tikzcd}$$

We start with a pair of functions at top right and proceed in two ways:

- left: anti-differentiate them, then go down: add the results; or
- down: add them, then go left: anti-differentiate the results.

The result is the same!

When the functions are specific, a brute force approach might be best:

Theorem 2.15.4: Componentwise Differentiation

The components of the derivative are the derivatives of the components:

$$\boxed{< f_1(t), ..., f_n(t) >' = < f_1'(t), ..., f_n'(t) >}$$

Theorem 2.15.5: Componentwise Integration

The components of an antiderivative are the antiderivatives of the components:

$$\boxed{\int < f_1(t), ..., f_n(t) > \, dt = < \int f_1(t)\, dt, ..., \int f_n(t)\, dt >}$$

Example 2.15.6: circular velocity

Let's integrate:

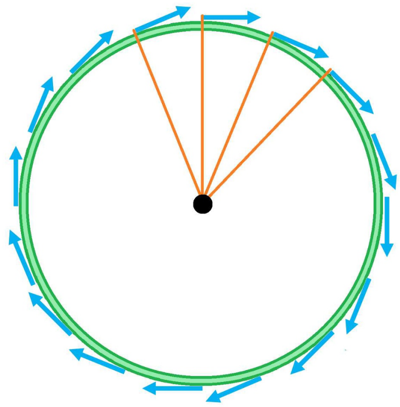
$$\begin{aligned} \int < \cos t,\ \sin t > \, dt &= < \int \cos t\, dt,\ \int \sin t\, dt > \\ &= < \sin t + C,\ \cos t + K > \\ &= < \sin t,\ \cos t > + < C, K > . \end{aligned}$$

Here C and K are just arbitrary constants which makes $< C, K >$ just an arbitrary vector.

2.16. The speed

Driving at a constant speed allows one to appreciate the scenery... and to study the shape of the road!

We already know that if we are moving in such a way that the distance to the origin remains the same, then *our velocity is always perpendicular to the vector of location.*



For the discrete case:

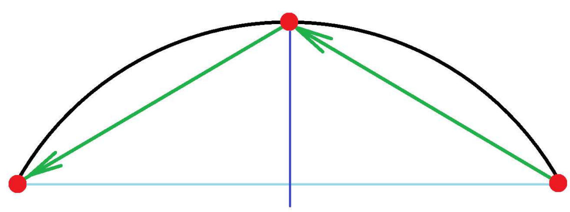
$$||F(t)|| = r \implies \frac{\Delta F}{\Delta t}(t) \cdot \Big(F(t + \Delta t) + F(t) \Big) = 0 \,.$$

This means that the difference quotient $\frac{\Delta F}{\Delta t}$ is perpendicular to the average of the two consecutive locations. For the continuous case:

$$||F(t)|| = r \implies F' \cdot F = 0 \,.$$

What about the relation between *the velocity and the acceleration*?

Suppose the difference quotient has a constant magnitude:



Then we make an important geometric observation: The median of an isosceles is also its height.

Therefore, we conclude:

- ▶ The velocity (over a double interval of time) is perpendicular to the acceleration.

Now, the derivatives. We see the following as the velocity and the acceleration of the motion along the circle:

$$F'(t) = < -\sin t, \ \cos t > \quad \text{and} \quad F''(t) = < -\cos t, \ -\sin t > \ .$$

Therefore:

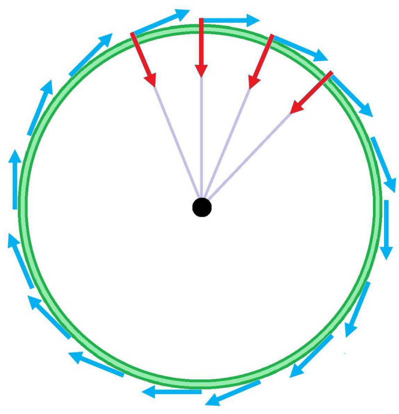
$$F'(t) \perp F''(t) \,.$$

For the general case, we suppose that our motion given by $X = F(t)$ is conducted in such a way that the *speed* remains the same:

$$||F'(t)|| = s \,.$$

Then,

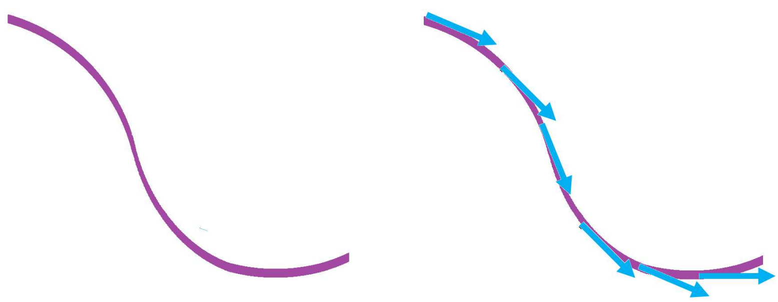
$$\frac{d}{dt}||F'||^2 = 0 \implies \frac{d}{dt}\big(F' \cdot F'\big) = 0 \implies F'' \cdot F' + F' \cdot F'' = 0 \implies F'' \cdot F' = 0 \,.$$



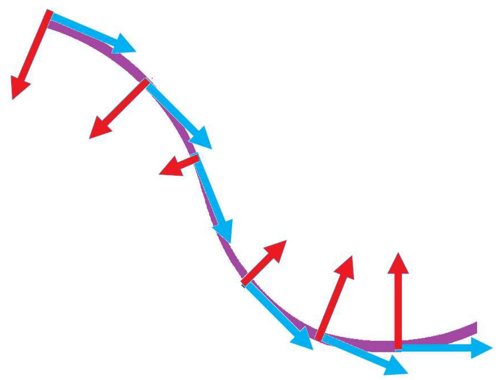
So, we have:

- ▶ *The acceleration is always perpendicular to the velocity.*

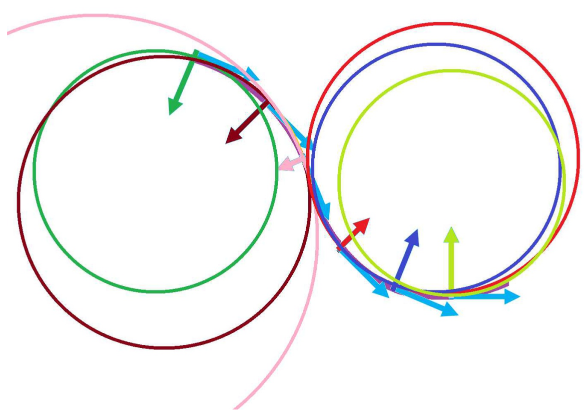
The advantage of this reformulation is that is applicable to *all dimensions*:



Just because the magnitude of the velocity is constant doesn't mean that the magnitude of the acceleration is zero (or even constant). The latter will depend on the sharpness of the turn, i.e., the curvature of the curve:



If we compare the result to the example we started with, we realize that it is as if – at every moment of time – we are moving along a circle at a constant speed and the acceleration points to the center of this circle... but the circles are constantly changing:



We now realize that not only the relation between two pairs of parametric curves is the same (a function and its derivative) but also that the two computations are exactly the same!

Theorem 2.16.1: Discrete Curve on Sphere

For any non-constant parametric curve $U = G(t)$ *defined at the nodes of a partition, we have:*

$$||G|| = r > 0 \implies \frac{\Delta G}{\Delta t}(c) \perp \frac{G(t) + G(t + \Delta t)}{2},$$

where c *is the corresponding secondary node, provided* $\Delta G(c) \neq 0$.

Theorem 2.16.2: Curve on Sphere

For any differentiable parametric curve $U = G(t)$, *we have:*

$$||G|| = r > 0 \implies \frac{dG}{dt} \perp G,$$

provided $\frac{dG}{dt} \neq 0$.

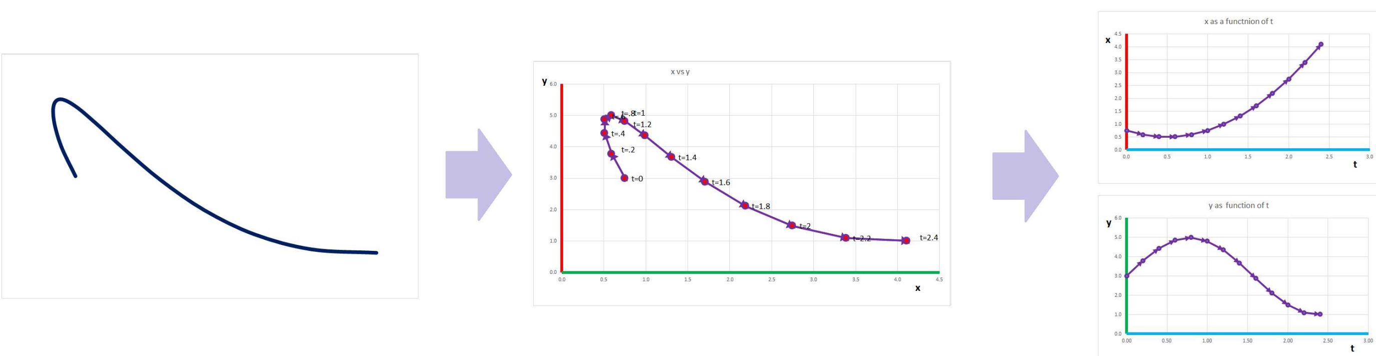
Driving at a constant speed, we never feel pressed into our seat or the seat belt but only feel the side-to-side swing. A constant speed parametrization takes the magnitude of the velocity – how fast we progress forward – out of consideration and allows us to concentrate on the curvature – how fast the direction is changing. They will be a major tool of our study of the shapes of curves.

Curves vs. parametric curves: we are now transitioning from studying driving to studying roads.

In mathematical terms, we transition from parametric curves to curves. But what is a curve?

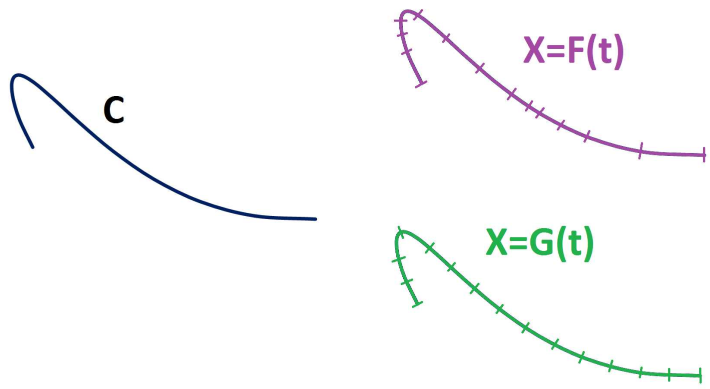
Definition 2.16.3: parametrization of curve

A *curve* C in $\mathbf{R}^n$ is the path of a continuous parametric curve, called its *parametrization*.



Then *a curve is the combination of all of its parametrizations* and we will need to answer the question: what do all these parametrizations have in common?

Suppose we have two parametrizations $X = G(s)$ and $X = F(t)$ of curve C. How are they related to each other?



A *change of variables* is a function used for substitution:

$$t = g(s)\,,$$

that turns one into the other:

$$G(s) = F(g(s))\,.$$

This composition is seen in this commutative diagram:

$$\begin{array}{ccc} [a,b] & \xrightarrow{F} & \mathbf{R}^n \\ \uparrow g & \nearrow G & \\ [c,d] & & \end{array}$$

Then, if F is a parametrization of C, G is called a *re-parametrization*.

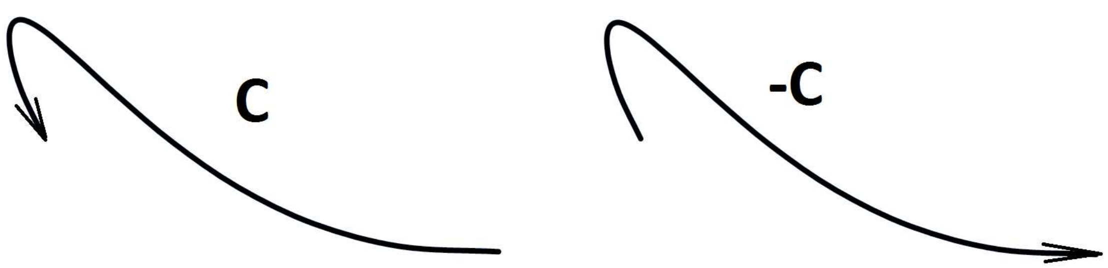
We know that as long as the function $t = g(s)$ of the change of variables is *one-to-one and onto*, the new path will be the same as the old. It's as if a driver made a recording of his drive along the road and we just run it at a different (possible variable) speed. However, this is only possible when the driver didn't stop or turned around.

Definition 2.16.4: regular parametrization

Suppose $X = F(t)$ is a parametric curve defined on $[a, b]$. If
- F is differentiable on (a, b),
- F' is continuous on (a, b), and
- $F' \neq 0$ on (a, b),

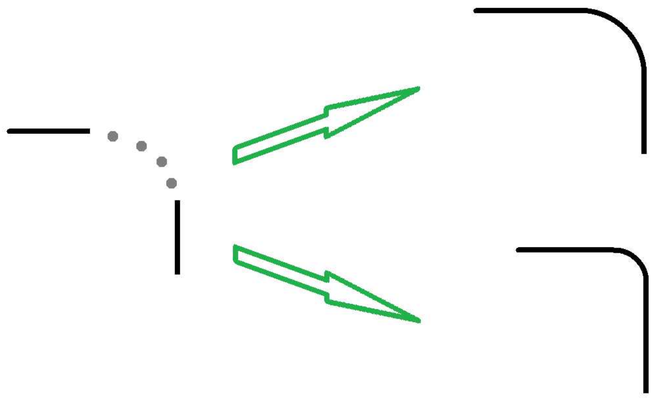
the curve is called *regular*. It is called a *regular parametrization* of the path.

Now, what if the two drivers drove in the *opposite* directions?



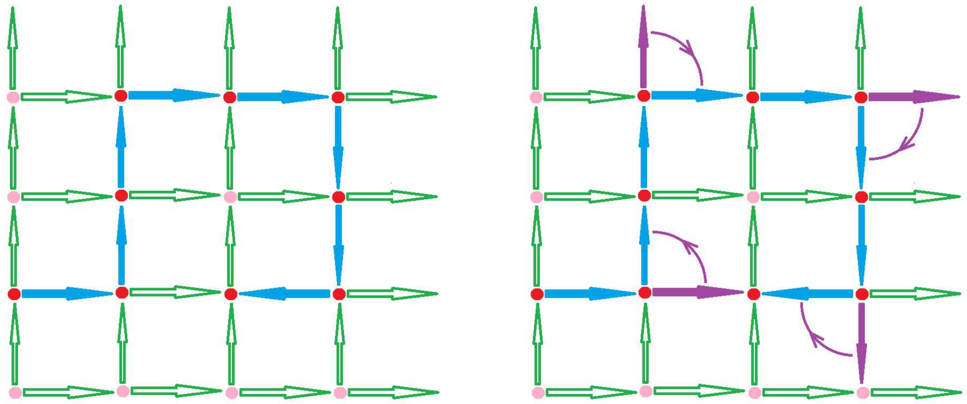
2.17. The curvature

We would like to devise a way to evaluate sharpness of turns of the road:



Since, depending on the car, we may need smaller or larger turning circles, we will know if this is possible.

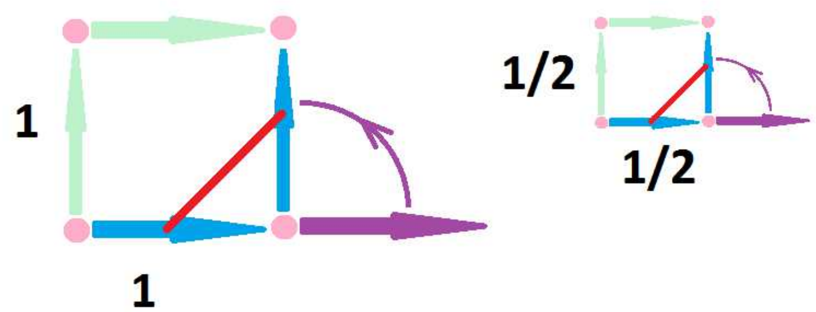
If we progress incrementally, the curvature is simply the turn that we make at each step. At its simplest, the road is a city street. We then move from intersection to intersection.



These are the only options:

turn	rotation	
no turn	0 degrees	0
left turn	90 degrees	$\pi/2$
right turn	90 degrees	$\pi/2$
U-turn	180 degrees	π

How do we find the angle? We copy each vector, attach the copy to its end point, and compare to the next vector. Does this mean that the curvature a 90-degree turn is $\pi/2$? No, because, as we know, the curvature scales down as the curve is scaled up, proportionally.



In order to measure the curvature in a scale independent manner, we use the angle of turn *per* distance covered, specifically, the distance between their centers of the two edges. For example, on the standard grid with edges of length 1, this distance is $\frac{\sqrt{2}}{2}$ and, therefore, the curvature of a 90-degree turn is $\frac{\pi}{\sqrt{2}}$. It is when the *diagonal* of the grid is 1, we have the curvature of such a turn equal to $\pi/2$.

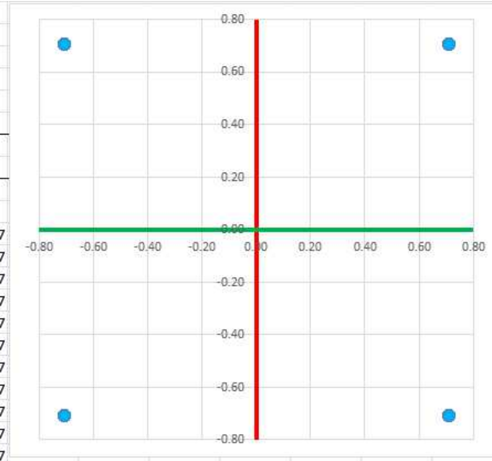
Example 2.17.1: square and circle

Let's confirm that this approach makes sense by considering a square and circle. We plot the unit circle,

$$F(t) =< \cos t,\ \sin t >,$$

with n points. For $n = 4$, we have a square with a unit diagonal.

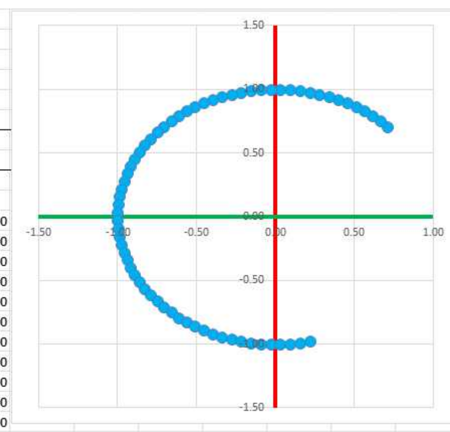
			interval	0	6.28						
			n=		4						
h											
1.57		1.00	1.00	feet	feet	feet	1	1	rad	feet	1/feet
k				displacement		magnitud	unit vector		angle		curv
0	t	x	y	x	y		x	y		d	
1	0.79	0.71	0.71								
2	2.36	-0.71	0.71	-1.41	0.00	1.41	-1.00	0.00			
3	3.93	-0.71	-0.71	0.00	-1.41	1.41	0.00	-1.00	1.57	1.00	1.57
4	5.50	0.71	-0.71	1.41	0.00	1.41	1.00	0.00	1.57	1.00	1.57
5	7.07	0.71	0.71	0.00	1.41	1.41	0.00	1.00	1.57	1.00	1.57
6	8.64	-0.71	0.71	-1.41	0.00	1.41	-1.00	0.00	1.57	1.00	1.57
7	10.21	-0.71	-0.71	0.00	-1.41	1.41	0.00	-1.00	1.57	1.00	1.57
8	11.78	0.71	-0.71	1.41	0.00	1.41	1.00	0.00	1.57	1.00	1.57
9	13.35	0.71	0.71	0.00	1.41	1.41	0.00	1.00	1.57	1.00	1.57
10	14.92	-0.71	0.71	-1.41	0.00	1.41	-1.00	0.00	1.57	1.00	1.57
11	16.49	-0.71	-0.71	0.00	-1.41	1.41	0.00	-1.00	1.57	1.00	1.57
12	18.06	0.71	-0.71	1.41	0.00	1.41	1.00	0.00	1.57	1.00	1.57
13	19.63	0.71	0.71	0.00	1.41	1.41	0.00	1.00	1.57	1.00	1.57



Here, we see the following columns. We compute the locations, then displacement for each two locations, then normalize them, then compute the angle between them (as the arccosine of their dot product), then finally divide by the half of the magnitude of the difference of the two displacements. The result is that the curvature is $\pi/2$.

Now, for $n = 100$, we have something very close to a unit circle. Through the same computations, the curvature is 1:

			interval	0	6.28						
			n=		100						
h											
0.06		1.00	1.00	feet	feet	feet	1	1	rad	feet	1/feet
k				displacement		magnitud	unit vector		angle		curv
0	t	x	y	x	y		x	y		d	
1	0.79	0.71	0.71								
2	0.85	0.66	0.75	-0.05	0.04	0.06	-0.73	0.68			
3	0.91	0.61	0.79	-0.05	0.04	0.06	-0.77	0.64	0.06	0.06	1.00
4	0.97	0.56	0.83	-0.05	0.04	0.06	-0.81	0.59	0.06	0.06	1.00
5	1.04	0.51	0.86	-0.05	0.03	0.06	-0.84	0.54	0.06	0.06	1.00
6	1.10	0.45	0.89	-0.06	0.03	0.06	-0.88	0.48	0.06	0.06	1.00
7	1.16	0.40	0.92	-0.06	0.03	0.06	-0.90	0.43	0.06	0.06	1.00
8	1.23	0.34	0.94	-0.06	0.02	0.06	-0.93	0.37	0.06	0.06	1.00
9	1.29	0.28	0.96	-0.06	0.02	0.06	-0.95	0.31	0.06	0.06	1.00
10	1.35	0.22	0.98	-0.06	0.02	0.06	-0.97	0.25	0.06	0.06	1.00
11	1.41	0.16	0.99	-0.06	0.01	0.06	-0.98	0.19	0.06	0.06	1.00
12	1.48	0.09	1.00	-0.06	0.01	0.06	-0.99	0.13	0.06	0.06	1.00
13	1.54	0.03	1.00	-0.06	0.00	0.06	-1.00	0.06	0.06	0.06	1.00



This means that we have to turn π radians per π units of lengths traveled. Changing the radius from 1 to 2 makes the curvature become $1/2$.

Now, suppose we have three distinct points P_1, P_2, P_3 in $\mathbf{R}^n$. What is the meaning of the curvature of this "curve"? It's not just its angle at P_2, i.e., $\widehat{P_1P_2P_3}$, but the angle the vectors P_1P_2 and P_1P_2 divided by the distance between their centers. We can choose to measure the turn not as an angle but rather the difference of the two vectors.

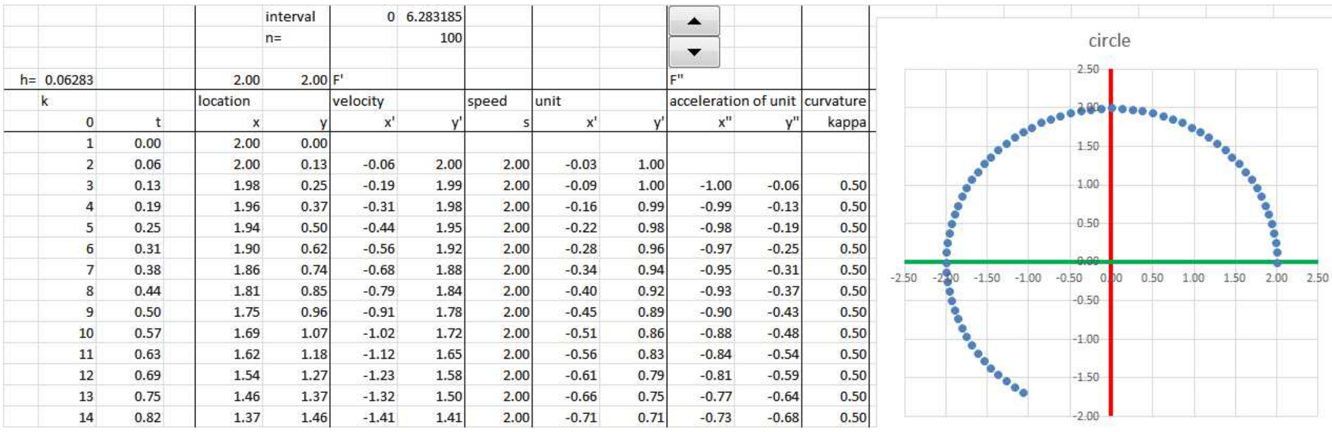
Suppose a sequence C of distinct points in $\mathbf{R}^n$ is represented by a parametric curve $X_k = F(t_k)$ defined at the nodes of a partition. Let's try this formula for the curvature of a parametric curve:

$$\varkappa(c_k) = \left| \frac{\big(F(t_k) - F(t_{k-1})\big) - \big(F(t_{k+1}) - F(t_k)\big)}{(F(t_{k+1}) - F(t_{k-1}))/2} \right|.$$

Example 2.17.2: circle

We apply the formulas to the standard parametrization:

$$F(t) =< 2\cos t,\ 2\sin t > .$$



			interval	0	6.283185						
			n=		100						
h= 0.06283		2.00	2.00	F'					F''		
k		location		velocity		speed	unit		acceleration of unit		curvature
0	t	x	y	x'	y'	s	x'	y'	x''	y''	kappa
1	0.00	2.00	0.00								
2	0.06	2.00	0.13	-0.06	2.00	2.00	-0.03	1.00			
3	0.13	1.98	0.25	-0.19	1.99	2.00	-0.09	1.00	-1.00	-0.06	0.50
4	0.19	1.96	0.37	-0.31	1.98	2.00	-0.16	0.99	-0.99	-0.13	0.50
5	0.25	1.94	0.50	-0.44	1.95	2.00	-0.22	0.98	-0.98	-0.19	0.50
6	0.31	1.90	0.62	-0.56	1.92	2.00	-0.28	0.96	-0.97	-0.25	0.50
7	0.38	1.86	0.74	-0.68	1.88	2.00	-0.34	0.94	-0.95	-0.31	0.50
8	0.44	1.81	0.85	-0.79	1.84	2.00	-0.40	0.92	-0.93	-0.37	0.50
9	0.50	1.75	0.96	-0.91	1.78	2.00	-0.45	0.89	-0.90	-0.43	0.50
10	0.57	1.69	1.07	-1.02	1.72	2.00	-0.51	0.86	-0.88	-0.48	0.50
11	0.63	1.62	1.18	-1.12	1.65	2.00	-0.56	0.83	-0.84	-0.54	0.50
12	0.69	1.54	1.27	-1.23	1.58	2.00	-0.61	0.79	-0.81	-0.59	0.50
13	0.75	1.46	1.37	-1.32	1.50	2.00	-0.66	0.75	-0.77	-0.64	0.50
14	0.82	1.37	1.46	-1.41	1.41	2.00	-0.71	0.71	-0.73	-0.68	0.50

The curvature is $1/2$ as expected.

We recognize some of these computations as difference quotients... Let's try this formula for the curvature of a parametric curve:

$$\varkappa = \left| \frac{\Delta}{\Delta t}\left(\frac{\Delta T}{\Delta t}\right) \right| .$$

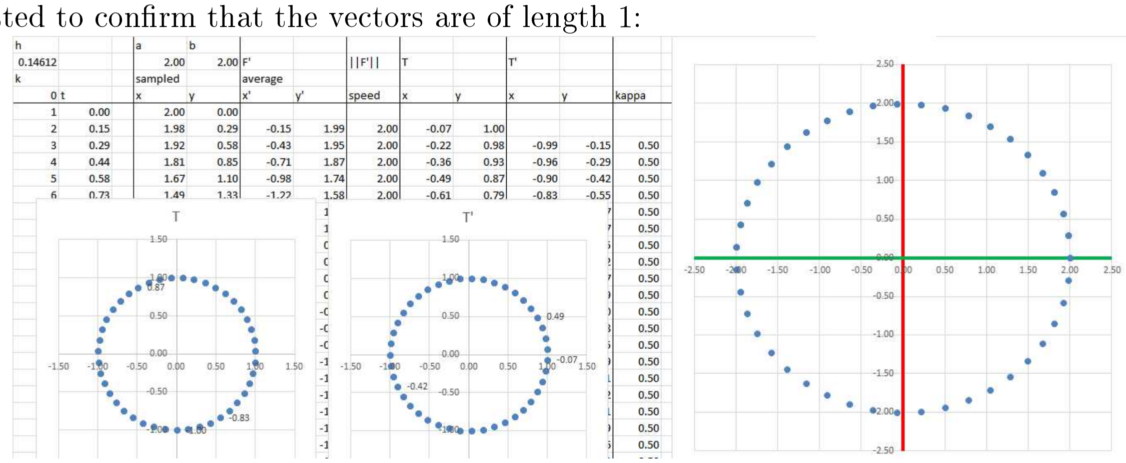
where T is the normalized vector of the velocity.

Example 2.17.3: circle

The curve $X = F(t)$ is sampled in the two columns marked x and y and then use the difference quotient to get the average velocity $\frac{\Delta F}{\Delta t}$ in the columns marked x' and y'. This time, we compute the speed s as the magnitude of this vector and then *normalize* the velocity to get the unit tangent vector T:

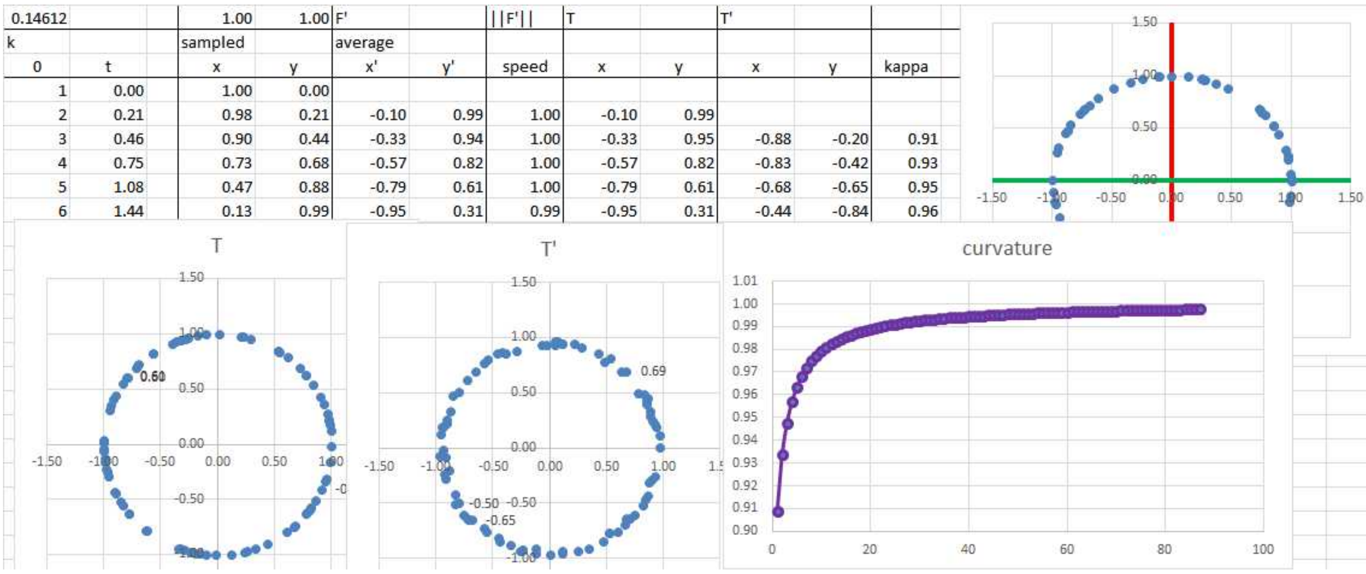
```
=RC[-3]/RC8
```

It is plotted to confirm that the vectors are of length 1:



h		a	b								
0.14612		2.00	2.00	F'		||F'||	T		T'		
k		sampled		average							
0	t	x	y	x'	y'	speed	x	y	x	y	kappa
1	0.00	2.00	0.00								
2	0.15	1.98	0.29	-0.15	1.99	2.00	-0.07	1.00			
3	0.29	1.92	0.58	-0.43	1.95	2.00	-0.22	0.98	-0.99	-0.15	0.50
4	0.44	1.81	0.85	-0.71	1.87	2.00	-0.36	0.93	-0.96	-0.29	0.50
5	0.58	1.67	1.10	-0.98	1.74	2.00	-0.49	0.87	-0.90	-0.42	0.50
6	0.73	1.49	1.33	-1.22	1.58	2.00	-0.61	0.79	-0.83	-0.55	0.50

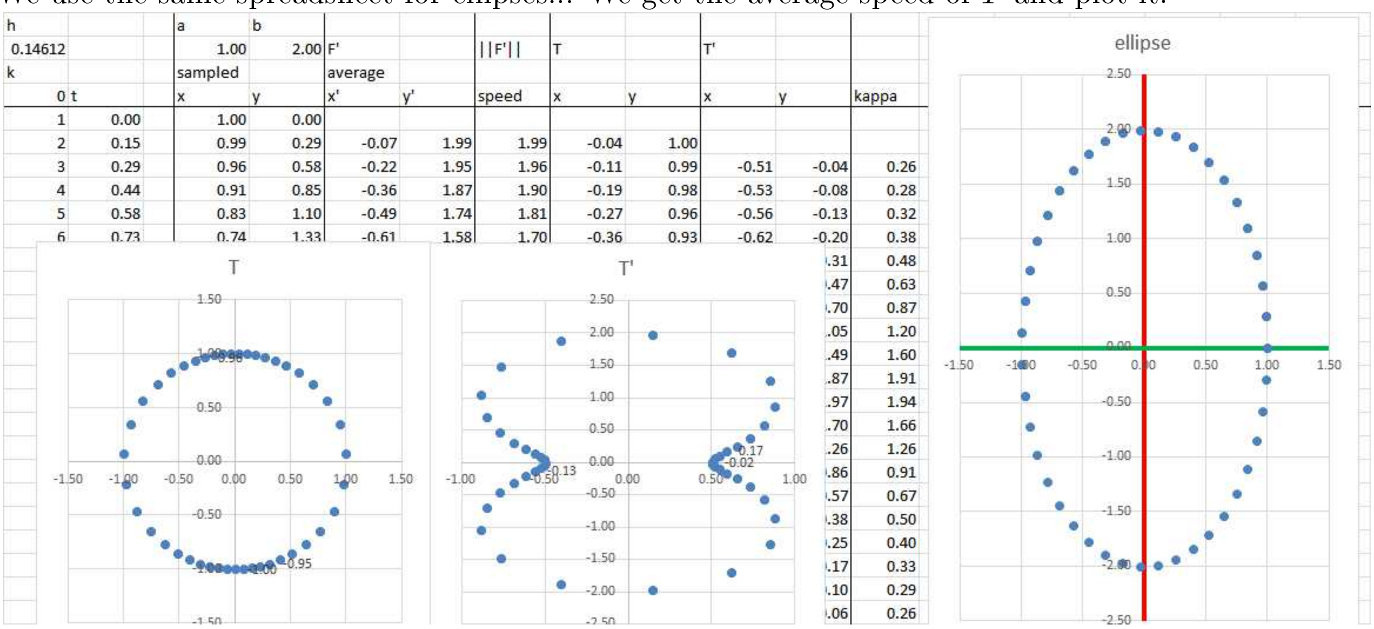
We use the difference quotient again to get the average rate of change of T and put it into the columns marked $T'x$ and $T'y$. It is also plotted. Finally, the magnitude of this vector is computed divided by the speed s; that's the curvature. It is constant at $1/2$ as expected. A parametrization with a non-constant speed produces the same result:



0.14612		1.00	1.00	F'		||F'||	T		T'		
k		sampled		average							
0	t	x	y	x'	y'	speed	x	y	x	y	kappa
1	0.00	1.00	0.00								
2	0.21	0.98	0.21	-0.10	0.99	1.00	-0.10	0.99			
3	0.46	0.90	0.44	-0.33	0.94	1.00	-0.33	0.95	-0.88	-0.20	0.91
4	0.75	0.73	0.68	-0.57	0.82	1.00	-0.57	0.82	-0.83	-0.42	0.93
5	1.08	0.47	0.88	-0.79	0.61	1.00	-0.79	0.61	-0.68	-0.65	0.95
6	1.44	0.13	0.99	-0.95	0.31	0.99	-0.95	0.31	-0.44	-0.84	0.96

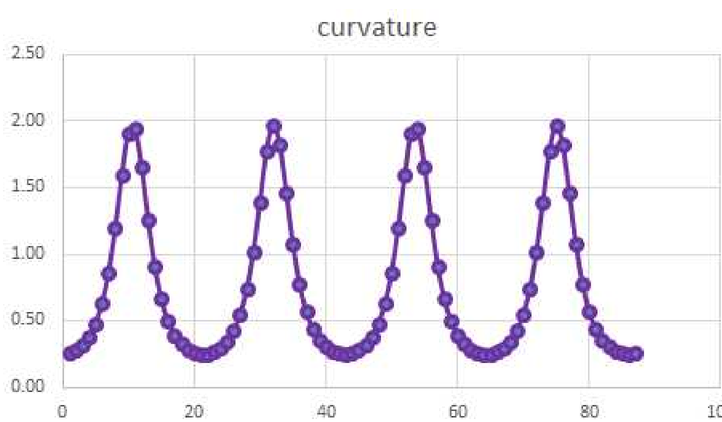
Example 2.17.4: ellipse

We use the same spreadsheet for ellipses... We get the average speed of T and plot it:

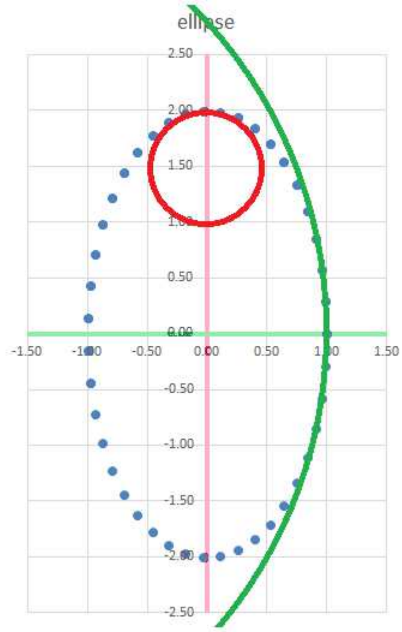


h		a	b								
0.14612		1.00	2.00	F'		||F'||	T		T'		
k		sampled		average							
0	t	x	y	x'	y'	speed	x	y	x	y	kappa
1	0.00	1.00	0.00								
2	0.15	0.99	0.29	-0.07	1.99	1.99	-0.04	1.00			
3	0.29	0.96	0.58	-0.22	1.95	1.96	-0.11	0.99	-0.51	-0.04	0.26
4	0.44	0.91	0.85	-0.36	1.87	1.90	-0.19	0.98	-0.53	-0.08	0.28
5	0.58	0.83	1.10	-0.49	1.74	1.81	-0.27	0.96	-0.56	-0.13	0.32
6	0.73	0.74	1.33	-0.61	1.58	1.70	-0.36	0.93	-0.62	-0.20	0.38
										.31	0.48
										.47	0.63
										.70	0.87
										.05	1.20
										.49	1.60
										.87	1.91
										.97	1.94
										.70	1.66
										.26	1.26
										.86	0.91
										.57	0.67
										.38	0.50
										.25	0.40
										.17	0.33
										.10	0.29
										.06	0.26

We recognize that at the start – at $(1, 0)$ – we have a vertical tangent vector and then it turns counterclockwise. The magnitude of this vector is computed divided by the speed s; that's the curvature. It is also plotted:



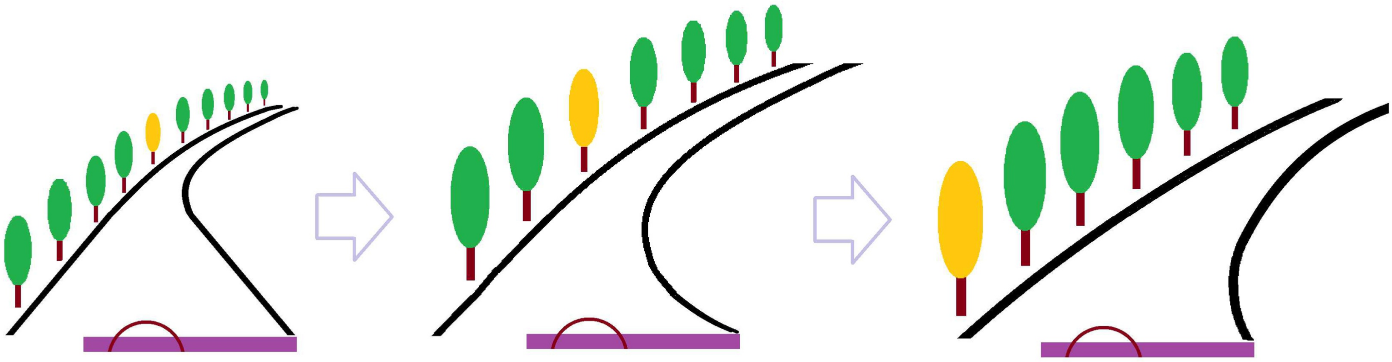
We can see how it is low at first at the flatter end of the ellipse and then grows to a larger value as we reach the sharper end. It runs between $1/4$ and 2. Two circles with radii 4 and $1/2$ are seen to approximate the ellipse at its two ends:



2.18. The arc-length parametrization

The main interest about these roads is their *curvature*: Is the turn too sharp so that the motorists may have trouble to staying on the road at a particular speed? How can we find out? There is no other way to learn the shape of the road but to drive over it! We need drivers, but what kinds of drivers are best for our purposes?

Imagine yourself driving. At first the road is straight which is recognized by the fact that you see a tree ahead and continue to see it to remain straight ahead. Then, as you start to turn, the trees start to pass your field of vision from right to left:



The trees may pass faster or slower. When it is faster, does it mean that the curvature of the road is higher? Not necessarily! Maybe you are just driving faster... So, we don't want drivers who drive erratically: speed up and slow down, or stop, or even turn around. The perfect driver for the job drives at a *constant speed*!

Example 2.18.1: circle

The standard parametrization of the circle of radius 2 is:

$$F(t) = < 2\cos t,\ 2\sin t > \ .$$

However, we cover the whole circle – length 4π – in 2π seconds. Let's slow down – by a factor of 2 –

and drive with speed 1! This is how:

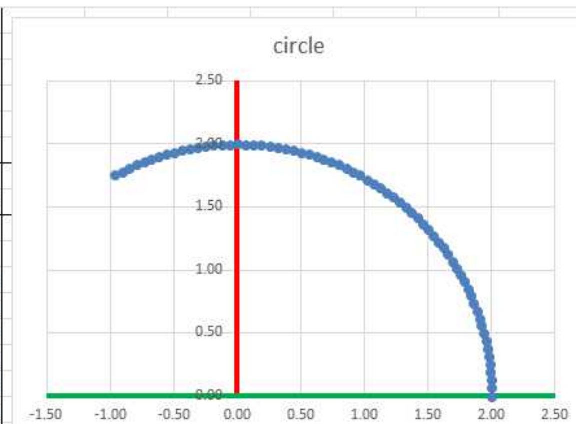
$$G(t) =< 2\cos(t/2),\ 2\sin(t/2) > \ .$$

Then in the formula for the curvature of a parametric curve discussed in the last section the normalization step disappears and the two difference quotients merge into the second difference quotient:

$$\varkappa = \left|\left|\frac{\Delta^2 G}{\Delta t^2}\right|\right| \ .$$

For computations, we apply the difference quotient twice and then find the magnitude of the resulting vector:

			interval	0	6.28319			
			n=		100			
h								
0.06283		2.00	2.00	F'		F"		
k		sampled		DQ		DQ		
0	t	x	y	x'	y'	x"	y"	kappa
1	0.00	2.00	0.00					
2	0.06	2.00	0.06	-0.02	1.00			
3	0.13	2.00	0.13	-0.05	1.00	-0.50	-0.02	0.50
4	0.19	1.99	0.19	-0.08	1.00	-0.50	-0.03	0.50
5	0.25	1.98	0.25	-0.11	0.99	-0.50	-0.05	0.50
6	0.31	1.98	0.31	-0.14	0.99	-0.50	-0.06	0.50
7	0.38	1.96	0.37	-0.17	0.99	-0.49	-0.08	0.50
8	0.44	1.95	0.44	-0.20	0.98	-0.49	-0.09	0.50



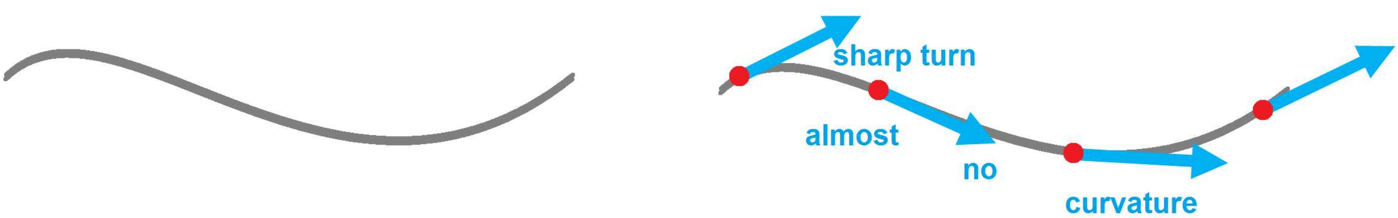
The curvature is 1/2 as expected.

At its simplest, the motion is described by the following two concepts:

velocity	$= \dfrac{\text{displacement}}{\text{time}}$		vectors	parametric curves
speed	$= \dfrac{\text{distance}}{\text{time}}$	$= \dfrac{\|\|\text{displacement}\|\|}{\text{time}}$	numbers	numerical functions

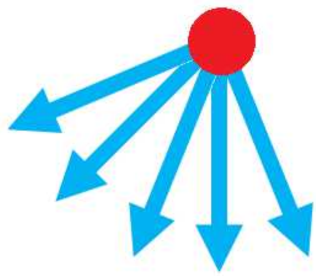
The familiar way to look at the curvature is to measure how much the curve deviates from a straight line. But what line is that? There is no single line.

At every location we will look at how much our motion takes us away from the current *tangent line*:



The curvature is then a measure of how fast the direction is changing. It is the measure of *how fast the tangent line is turning.*

Of course, this is not the same as to say that the curvature is how fast the tangent vector, F', is changing, which is F'', as this would make the concept dependent on the speed, $||F'||$. Unless, of course, the speed is fixed at, say, 1. In that case, we are measuring the speed of turning of a unit vector:



This speed is the same as the rate of change of this vector. Thus, for a "unit-speed" parametrization $X=G(s)$, i.e., $||G'||=1$, the curvature is the rate of change of the velocity vector G', i.e., the acceleration.

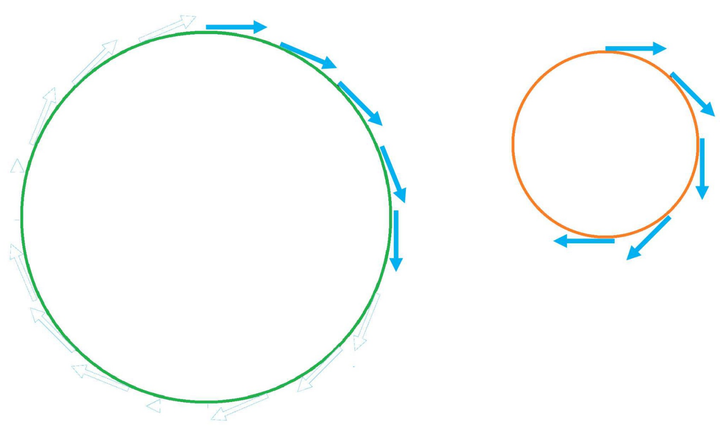
Definition 2.18.2: curvature

Suppose a parametrization $X=G(s)$ of a curve C satisfies $||G'||=1$, then the *curvature of the curve* is the numerical function that is the magnitude of the vector function of the second derivative:

$$\boxed{\varkappa(t)=||G''(t)||}$$

Example 2.18.3: curvatures of circles

Let's compare two circles of radius 2 and radius 1. Now, over the same distance, π, and, therefore, over the same time, covered by a point moving at speed 1 the rotation of the tangent is respectively $\pi/2$ and π:



The smaller circle has twice as large curvature! It seems that the curvature of a circle is the reciprocal of its radius... At the extreme, the large circle turns into a straight line with zero curvature!

Definition 2.18.4: arc-length parametrization

A parametrization $X=G(s)$ is called an *arc-length parametrization* if $||G'(s)||=1$.

Example 2.18.5: circle

The standard parametrization of the circle of radius R (a constant angular velocity),

$$F(t)=<R\cos t,\ R\sin t>,$$

isn't arc-length unless $R=1$ because

$$F'(t)=<-R\sin t,\ R\cos t> \Longrightarrow ||F'(t)||=\sqrt{(-R\sin t)^2+(R\cos t)^2}=R\sqrt{(\sin t)^2+(\cos t)^2}=R\,.$$

How can we make this into an arc-length parametrization? We have to be careful so that the modified curve is still the circle. The safest way to ensure that is to re-scale the time t. From the last example:

we have to move slower if $R < 1$ and faster if $R > 1$. What should we replace t with so that R is cancelled under differentiation? Just put:

$$t = s/R\,.$$

Then, indeed, we have:

$$G(s) =< R\cos(s/R),\ R\sin(s/R) > \Longrightarrow G'(s) =< -\sin(s/R),\ \cos(s/R) > \Longrightarrow ||G'(s)|| = 1\,,$$

by the *Pythagorean Theorem*.

Example 2.18.6: line

Let's consider this line:

$$F(t) =< 3, 4 > t \implies F'(t) =< 3, 4 > \Longrightarrow ||F'(t)|| = \sqrt{3^2 + 4^2} = 5\,.$$

How can we make this into an arc-length parametrization? We re-scale the time t again to slow it down:

$$t = s/5\,.$$

Thus, this is what we have found:

$$G(s) =< 3/5, 4/5 > s\,.$$

Here the direction vector is a unit vector!

Theorem 2.18.7: Constant Speed Parametrization

A parametric curve of a line,

$$F(t) = At + B\,,$$

is an arc-length parametrization if and only if $||A|| = 1$.

Thus, a well-chosen *change of variables*, i.e., a substitution:

$$t = g(s)\,,$$

can make a parametric curve into an arc-length parametrization:

$$G(s) = F(g(s))\,.$$

Since the function $t = g(s)$ of the change of variables is *one-to-one and onto*, the new parametric curve has the same path as the old. It's as if a random driver made a recording of his drive along the road we study, and now we would like to process the video in such a way that the drive appears to be at 1 mile per hour.

Such a new parameter s is called an *arc-length parameter* of the curve. There can be only two of those for a given curve – back and forth – the rest are just the shifts of those two.

In the first example above, these are:

$$t = s/R \ \text{ and } \ t = -s/R\,,$$

and in the second:

$$t = s/5 \ \text{ and } \ t = -s/5\,.$$

We can speak of *the* arc-length parametrization when the curve is oriented, i.e., its direction is indicated, and $t = g(s)$ is increasing.

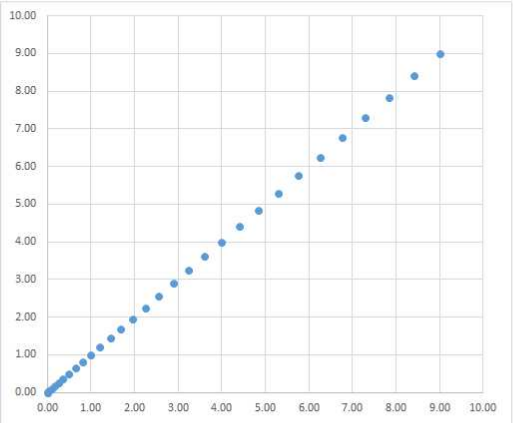
These simple (linear) changes of variables have worked only because the speed has been constant.

Example 2.18.8: straight acceleration

Suppose we are accelerating along the straight line $y = x$:

$$F(t) = < t^2,\ t^2 >,\ t \geq 0\,.$$

The denser dots indicate slower motion:



We will need a non-linear change of variables. We have:

$$F'(t) = < 2t, 2t > = < 1, 1 > 2t \implies ||F'(t)|| = \sqrt{2}t\,.$$

We would like to find such a function $t = g(s)$ that

$$G(s) = F(g(s)) \implies ||G'(s)|| = 1\,.$$

We substitute and use the *Chain Rule*:

$$||G'(s)|| = ||F'(g(s))g'(s)|| = ||< 1, 1 > 2g(s)g'(s)|| = 2\sqrt{2}g(s)g'(s) = 1\,.$$

Then,

$$g'(s) = \frac{1}{2\sqrt{2}g(s)}\,.$$

The answer isn't easy to guess, but here it is

$$g(s) = \sqrt{s/2}\,.$$

From the relation between the two parametrizations,

$$G(s) = F(g(s))\,,$$

we derive by the *Chain Rule*:

$$G' = F'g'\,.$$

The theorem below then follows.

Theorem 2.18.9: Arc-length Parametrization

Suppose $X = F(t)$ is a differentiable on an open interval parametric curve. Then, $t = g(s)$ is the change of variables that produces the arc-length parametrization $X = G(s) = F(g(s))$ if and only if

$$\boxed{g'(s) = \frac{1}{||F'(g(s))||}}$$

The non-zero derivative is then a prerequisite for arc-length re-parametrization.

Corollary 2.18.10: Length Is Integral of Speed

Suppose $X = F(t)$ is a regular parametric curve defined on $[a, b]$. Then,

$$\boxed{s = \int_a^t ||F'(u)||\, du, \ a \le t \le b}$$

is the arc-length parameter.

Exercise 2.18.11

Prove the corollary.

For each t, the integral is the length of the curve up to $F(t)$. The result is confirmed in the next section.

2.19. Re-parametrization

Let's recall that if a parametrization $X = G(s)$ of a curve C satisfies $||G'(s)|| = 1$, i.e., s is an arc-length parameter, the curvature of the curve was defined to be the magnitude of the vector function of the second derivative:

$$\varkappa(s) = ||G''(s)||\,.$$

Example 2.19.1: circle

For a circle of radius R, we use the *arc-length parametrization*:

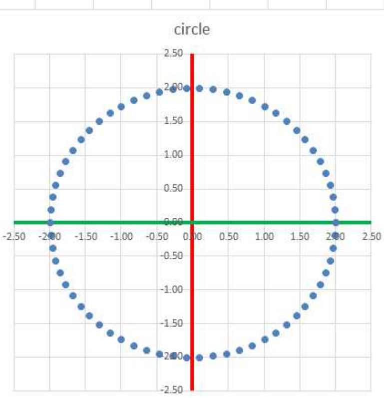
$$G(s) =< R\cos(s/R),\ \sin(s/R) > \ .$$

We start with $R = 2$. The curve is sampled in the two columns marked x and y and these points are plotted. We then use the difference quotient:

```
=(RC[-2]-R[-1]C[-2])/R6C1
```

We get the average velocity and put it into the columns marked x' and y'. We use the difference quotient again to get the average acceleration and put it into the columns marked x'' and y'':

h								
0.1904		2.00	2.00	F'		F''		
k		sampled		average				
0	t	x	y	x'	y'	x''	y''	kappa
1	0.00	2.00	0.00					
2	0.19	1.99	0.19	-0.05	1.00			
3	0.38	1.96	0.38	-0.14	0.99	-0.50	-0.05	0.50
4	0.57	1.92	0.56	-0.24	0.97	-0.49	-0.09	0.50
5	0.76	1.86	0.74	-0.33	0.94	-0.48	-0.14	0.50
6	0.95	1.78	0.92	-0.42	0.91	-0.46	-0.19	0.50
7	1.14	1.68	1.08	-0.50	0.87	-0.44	-0.23	0.50
8	1.33	1.57	1.24	-0.58	0.81	-0.42	-0.27	0.50
9	1.52	1.45	1.38	-0.65	0.76	-0.39	-0.31	0.50
10	1.71	1.31	1.51	-0.72	0.69	-0.36	-0.34	0.50
11	1.90	1.16	1.63	-0.79	0.62	-0.33	-0.38	0.50
12	2.09	1.00	1.73	-0.84	0.54	-0.29	-0.41	0.50
13	2.28	0.83	1.82	-0.89	0.46	-0.25	-0.43	0.50
14	2.48	0.65	1.89	-0.93	0.37	-0.21	-0.45	0.50
15	2.67	0.47	1.94	-0.96	0.28	-0.16	-0.47	0.50
16	2.86	0.28	1.98	-0.98	0.19	-0.12	-0.49	0.50
17	3.05	0.10	2.00	-1.00	0.10	-0.07	-0.49	0.50
18	3.24	-0.10	2.00	-1.00	0.00	-0.02	-0.50	0.50
19	3.43	-0.28	1.98	-1.00	-0.10	0.02	-0.50	0.50



Finally, the magnitude of this vector is computed; that's the curvature. The result is, as expected, 1/2.

Now we would like to try to express the curvature in terms of the original parameter t.

The curvature is the rate of change of the *direction* – given by the tangent line – of the curve. The derivative F' is tangent but it also changes is *magnitude*. To rectify that, we consider the *unit tangent vector* at the point $F(t)$:

$$T(t) = \frac{F'}{||F'||}\,.$$

This is an alternative way to state the definition.

Definition 2.19.2: curvature

Suppose $X = F(t)$ is a parametrization of a curve C with non-zero derivative, $F' \neq 0$. Then the *curvature* of the curve is defined to the speed of change of the unit tangent vector with respect to the arc-length parameter, i.e.,

$$\boxed{\varkappa = \left|\left|\frac{dT}{ds}\right|\right|}$$

Recall that our change of variables:

$$t \mapsto s\,,$$

creates an arc-length parametrization:

$$\begin{array}{ccc} t & \xrightarrow{F} & X \\ \updownarrow & \nearrow_G & \\ s & & \end{array}$$

As the change of variables $t \mapsto s$ is one-to-one and onto, it is *invertible*. That is why we can go along this arrow in either direction.

We will treat the diagram as a functional relation between three quantities. The rates of change of the three are then the familiar *related rates* from Volume 2 (Chapter 2DC-4). Just as above but in the Leibniz notation, we have by the *Chain Rule*:

$$\frac{dX}{dt} = \frac{dX}{ds}\frac{ds}{dt}\,.$$

We now use the fact that s is an arc-length parameter:

$$\left|\left|\frac{dX}{ds}\right|\right| = 1\,.$$

Then, we have:

$$\left|\left|\frac{dX}{dt}\right|\right| = \left|\left|\frac{dX}{ds}\right|\right| \left|\frac{ds}{dt}\right| = \left|\frac{ds}{dt}\right|\,,$$

Now assuming that s is increasing with respect to t (same direction!), we have a convenient way to describe the arc-length parameter:

$$\frac{ds}{dt} = ||F'||\,.$$

Its derivative is the *speed*! Naturally, if this is 1, this *is* the arc-length parameter, $s = t$.

Now, from the *Chain Rule* and the above identity, we have a formula for the derivative of T with respect to the *old* parameter:

$$\frac{dT}{dt} = \frac{dT}{ds}\frac{ds}{dt}\,.$$

Then,

$$\left|\left|\frac{dT}{dt}\right|\right| = \left|\left|\frac{dT}{ds}\right|\right| \frac{ds}{dt} = \left|\left|\frac{dT}{ds}\right|\right| ||F'|| = \varkappa\, ||F'||\,.$$

We have proven the following.

Theorem 2.19.3: Curvature

Suppose $X = F(t)$ is a parametrization of a curve C with non-zero derivative, $F' \neq 0$. Then the curvature of the curve is given by:

$$\boxed{\varkappa(t) = \frac{||T'(t)||}{||F'(t)||}}$$

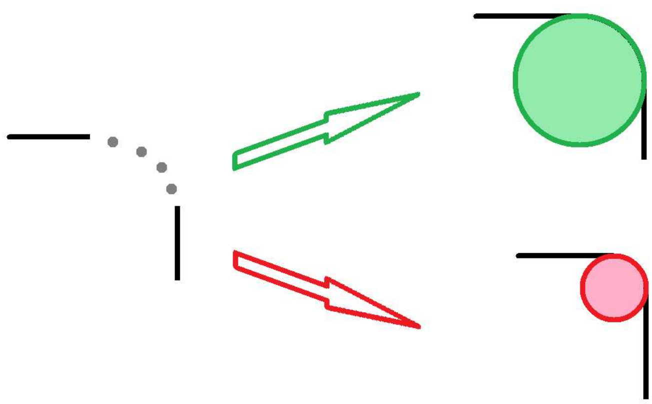
Definition 2.19.4: radius of curvature

The reciprocal of the curvature of a parametric curve $X = F(t)$ is called the *radius of curvature* of the curve:

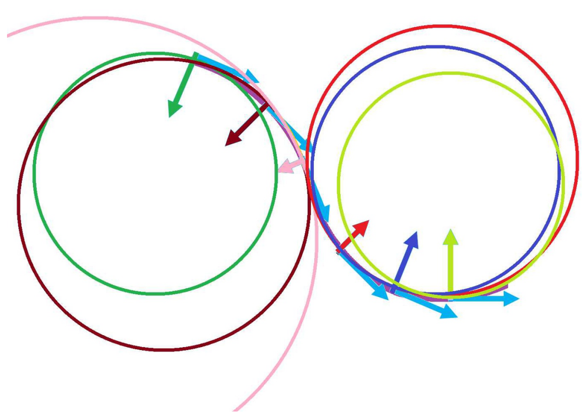
$$\boxed{R = \frac{1}{\varkappa}}$$

unless the former is 0, then the radius is said to be infinite. The *circle of curvature* at the point $Q = F(a)$ of the curve is the circle

- through this point,
- with its radius equal to the radius of the curvature of the curve, and
- with its center on the line perpendicular to the curve at Q.



In general, the circles of curvature are continuously evolving as we move along the curve:



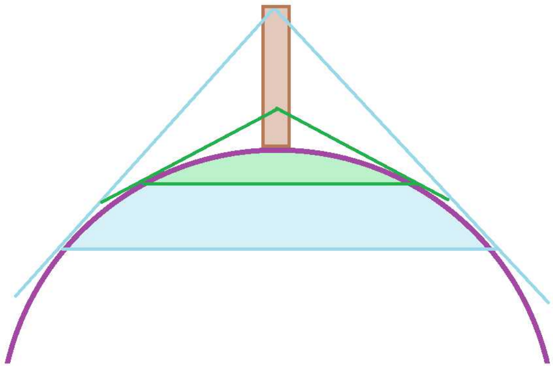
The curvature is an *intrinsic* property of the curve: Even if we are riding a car blind-folded, we can figure out the shape of the road from the direction and the strength of the pull. Indeed, the curvature is the acceleration and the latter feels like a force!

The job of finding the curvature is made easier if the car is driven at a constant, known speed. We don't need to walk the surrounding area and make measurements... In a similar manner, a bug traveling along a thin tube might be able to map its shape. Following this idea, we have to travel the universe to map its

curvature as we are unable to step outside and measure it.

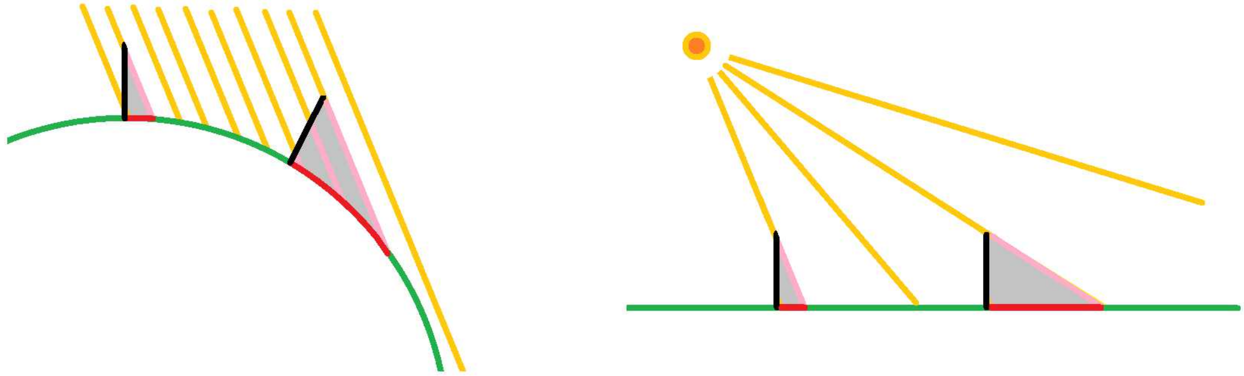
Example 2.19.5: Is the Earth flat?

How does one prove that the Earth is round? One known method is to climb a taller building and see that your field of vision has extended:



The same idea is to move away from a taller building and see that it disappears beyond the horizon. The challenge of this method is that one needs a perfectly flat (!?) area for the experiment.

Another method is to measure the shadows of two identical sticks in two different locations:



The method assumes however that the Sun is so large that its rays are essentially parallel.

All these methods have flaws but the main one is that they only prove that the Earth has curvature – at that particular location! You'd have to visit every location on the face of the Earth and repeat this experiment. Once you have confirmed that the curvature is the same everywhere, you'd still need some mathematics that proves roundness of the whole surface. But, if you've visited every place on Earth, why not make a map of it and prove its shape this way?!

Exercise 2.19.6

Show that if a plane curve has a constant curvature, it is either a straight line or a circle. What if the curve isn't necessarily plane?

2.20. Lengths of curves

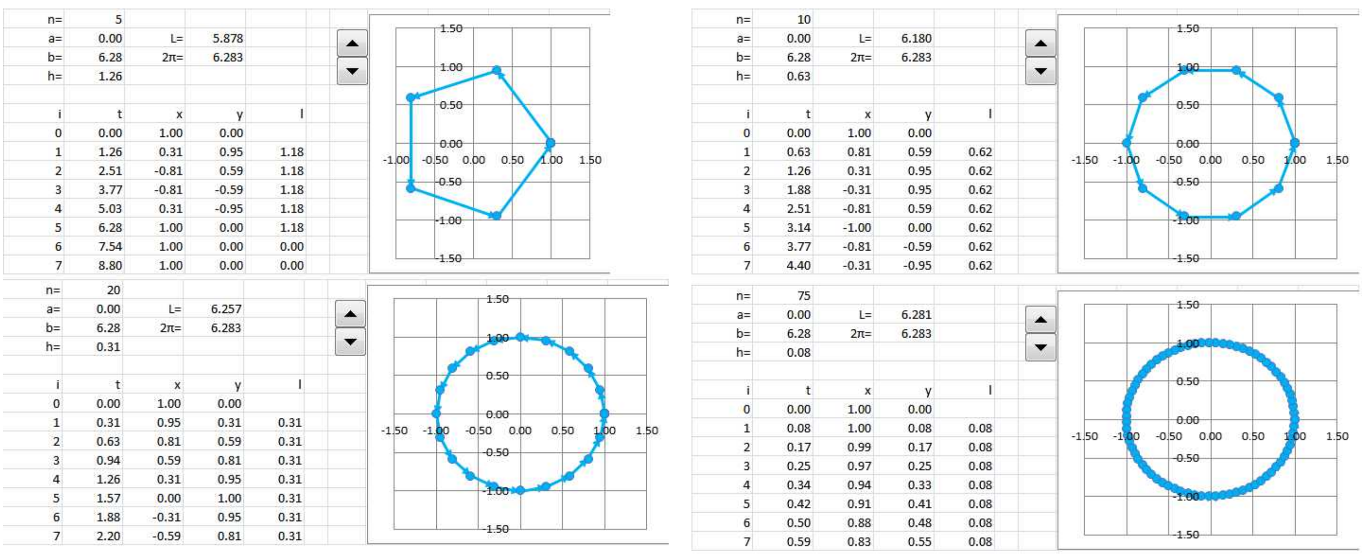
Example 2.20.1: circle

In Volume 3 (Chapter 3IC-1), we compute the length of the upper half of the unit circle as it is represented as the graph of a simple function. This time, let's compute length of the circle as a parametric curve. The idea is the same:

1. Place points on the curve.
2. Connect them consecutively by edges.
3. Approximate the curve with a continuous curve made of these edges.

The length of each edge is computed via the *Distance Formula*:

```
=SQRT((RC[-2]-R[-1]C[-2])^2+(RC[-1]-R[-1]C[-1])^2)
```



n=	5			
a=	0.00	L=	5.878	
b=	6.28	2π=	6.283	
h=	1.26			
i	t	x	y	l
0	0.00	1.00	0.00	
1	1.26	0.31	0.95	1.18
2	2.51	-0.81	0.59	1.18
3	3.77	-0.81	-0.59	1.18
4	5.03	0.31	-0.95	1.18
5	6.28	1.00	0.00	1.18
6	7.54	1.00	0.00	0.00
7	8.80	1.00	0.00	0.00

n=	10			
a=	0.00	L=	6.180	
b=	6.28	2π=	6.283	
h=	0.63			
i	t	x	y	l
0	0.00	1.00	0.00	
1	0.63	0.81	0.59	0.62
2	1.26	0.31	0.95	0.62
3	1.88	-0.31	0.95	0.62
4	2.51	-0.81	0.59	0.62
5	3.14	-1.00	0.00	0.62
6	3.77	-0.81	-0.59	0.62
7	4.40	-0.31	-0.95	0.62

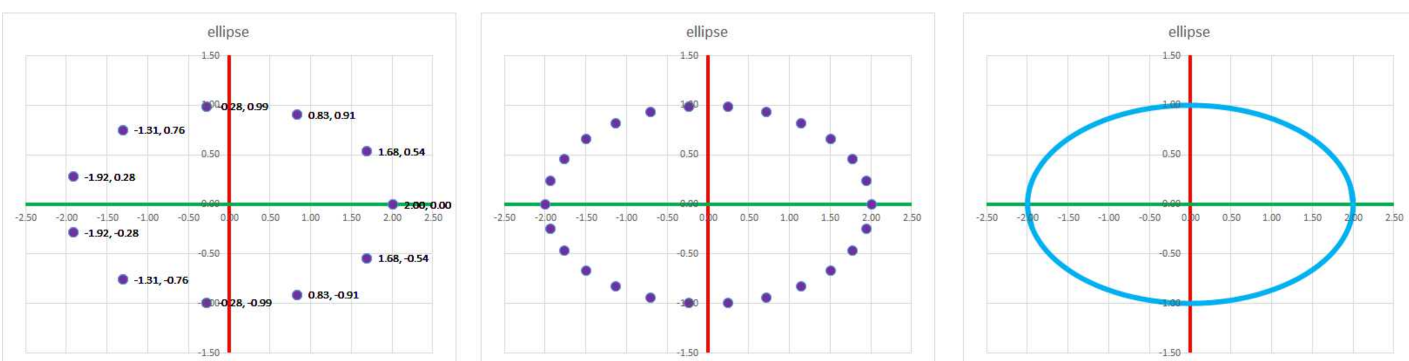
n=	20			
a=	0.00	L=	6.257	
b=	6.28	2π=	6.283	
h=	0.31			
i	t	x	y	l
0	0.00	1.00	0.00	
1	0.31	0.95	0.31	0.31
2	0.63	0.81	0.59	0.31
3	0.94	0.59	0.81	0.31
4	1.26	0.31	0.95	0.31
5	1.57	0.00	1.00	0.31
6	1.88	-0.31	0.95	0.31
7	2.20	-0.59	0.81	0.31

n=	75			
a=	0.00	L=	6.281	
b=	6.28	2π=	6.283	
h=	0.08			
i	t	x	y	l
0	0.00	1.00	0.00	
1	0.08	1.00	0.08	0.08
2	0.17	0.99	0.17	0.08
3	0.25	0.97	0.25	0.08
4	0.34	0.94	0.33	0.08
5	0.42	0.91	0.41	0.08
6	0.50	0.88	0.48	0.08
7	0.59	0.83	0.55	0.08

As we increase the number of segments, the result that we know to be correct, 2π, is being approached. Each of these curves is a 0-form over $\mathbf{R}$ with values in $\mathbf{R}^2$.

We now review and generalize the Riemann sum construction from Chapter 3IC-3 for computing the lengths of curves.

Suppose $X = F(t)$ is a parametric curve in the m-dimensional space defined at the nodes of a partition of an interval $[a, b]$. In other words, this is a 0-form over $\mathbf{R}$ with values in $\mathbf{R}^m$.



We have a partition of $[a, b]$ with n intervals. This is what happens to each interval $[t_{k-1}, t_k],\ k = 1, 2, ..., n$ of the partition: the curve leaps from $F(t_{k-1})$ to $F(t_k)$. The length of this segment is:

$$||F(t_k) - F(t_{k-1})||\,.$$

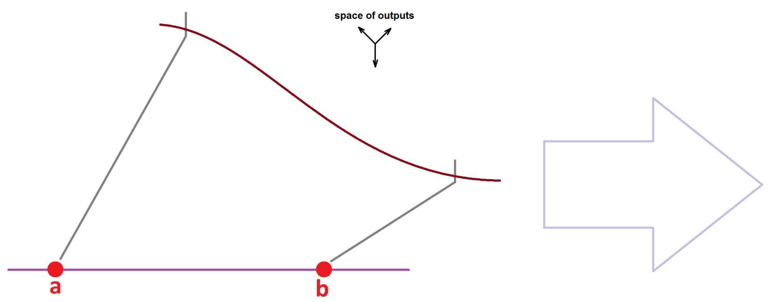
Thus, the full length of these segments is equal to the sum of all n of those, as follows:

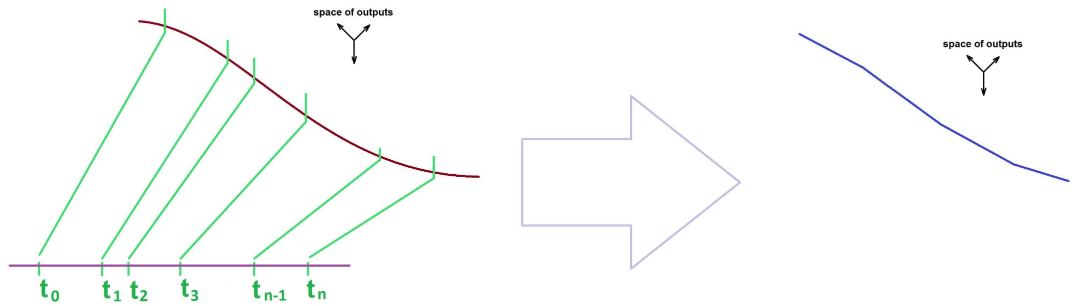
$$\text{total length } = \sum_{k=1}^{n} ||F(t_k) - F(t_{k-1})||\,.$$

This formula is to be used for approximations, just as in the last example. For exact answers, we make the intervals smaller and smaller. We also anticipate that this process will ends with a Riemann integral.

Suppose now that we have a parametric curve $X = F(t),\ a \le t \le b$. We will define and then compute the *length of the path of the parametric curve.*

We partition the interval to sample the curve and then we approximate the pieces of the curve with straight segments with the above formula.





Over each interval $[t_{k-1}, t_k],\ k = 1, 2, ..., n$ of the partition, the path of F goes from $F(t_{k-1})$ to $F(t_k)$. is replaced with a segment between these two points. Then, the full length of the path of F is approximated by:

$$\text{length of path } \approx \sum_{k=1}^{n} ||F(t_k) - F(t_{k-1})||\,.$$

Since this doesn't look like the Riemann sum of a function, we need to create the missing Δt in the formula. We also use the earlier insight: there must be the derivative of F present. We create the missing *difference quotient* by manipulating the formula.

The two goals match up: we divide and multiply each term by Δt, as follows:

$$\begin{aligned}\text{Sum of lengths } &= \sum_{k=1}^{n} ||F(t_k) - F(t_{k-1})||\\ &= \sum_{k=1}^{n} ||F(t_k) - F(t_{k-1})|| \cdot \frac{\Delta t}{\Delta t}\\ &= \sum_{k=1}^{n} \left|\left| \frac{1}{\Delta t}(F(t_k) - F(t_{k-1})) \right|\right| \cdot \Delta t\,.\end{aligned}$$

We then have both Δt and the difference quotient. But this is still not a Riemann sum; the expression that precedes Δt should be the value of some numerical function evaluated at the *secondary nodes* of the partition. We haven't specified those and this is the time to do that. We apply, as we've done before, the

Mean Value Theorem: There is some c_k in the interval $[t_{k-1}, t_k]$ such that

$$\frac{1}{\Delta t}(F(t_k) - F(t_{k-1})) = F'(c_k)\,.$$

Therefore,

$$\text{Sum of lengths } = \sum_{k=1}^{n} ||F'(c_k)||\,\Delta t\,.$$

This is the Riemann sum of the numerical function $g(t) = ||F'(t)||$ over the augmented partition of $[a, b]$ with the secondary nodes $c_1, ..., c_n$.

The analysis above reveals the *meaning* of the new concept.

Definition 2.20.2: length of curve

The *length of a curve* given by the path of a regular parametric curve $X = F(t)$ over interval $[a, b]$ is defined to be the integral

$$\boxed{L = \int_a^b ||F'(t)||\,dt}$$

if it exists.

Just as before, the function F itself is absent from the formula because only the shape (given by the derivative) and not its location matters for the length of the curve.

Theorem 2.20.3: Length of Curve

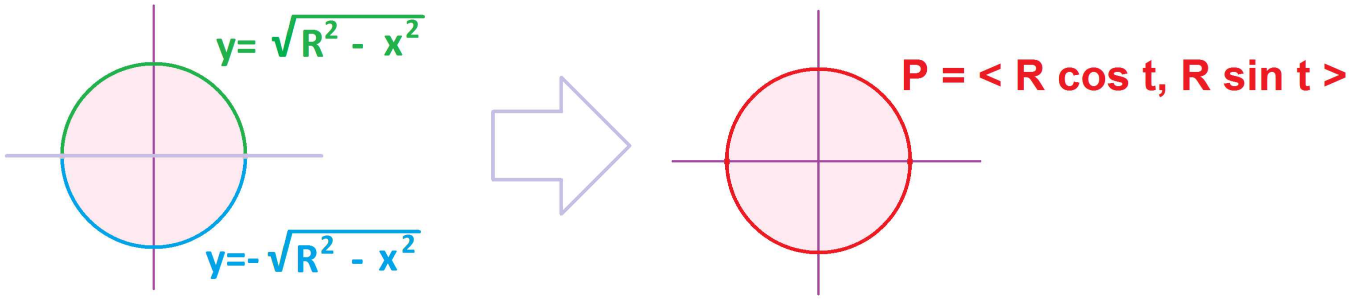
The length of a curve is defined and defined uniquely if it can be represented as the path of a regular parametric curve (i.e., it is parametrized by this function).

Proof.

We need the extra condition to ensure that the *Mean Value Theorem* applies and the resulting function is integrable.

Example 2.20.4: circumference

Let's re-prove that the circumference of a circle of radius R is $2\pi R$. We know the result from Volume 3 (Chapter 3IC-1), but this time we don't have to represent the curve as a combination of *two* graphs of functions.



We have a single parametric curve:

$$X = F(t) =< R\cos t,\ R\sin t >,\ 0 \le t \le 2\pi\,.$$

Then,

$$F'(x) =< -R\sin t,\ R\cos t >\ .$$

We apply the formula:

$$\begin{aligned}\text{The length} &= \int_0^{2\pi} ||F'||\,dt\\ &= \int_0^{2\pi} ||< -R\sin t,\ R\cos t >||\,dt\\ &= \int_0^{2\pi} R\,dt\\ &= 2\pi R\,.\end{aligned}$$

Much easier! Since originally we obtained this result via trig substitution, the new computation reveals its true meaning: parametrization.

Exercise 2.20.5

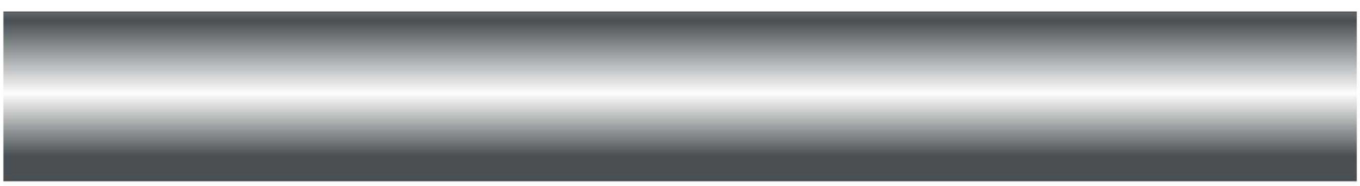
Find the length of the segment of the parabola $y = x^2$ from $(0,0)$ to $(1,1)$.

Exercise 2.20.6

Find the length of one arc of the cycloid: $x = R(1-\sin t),\ y = R(1-\cos t)$.

2.21. Arc-length integrals: weight

What if we need to find the *weight of a curve of variable density*? We know (from Volume 3, Chapter 3IC-3) the answer when this is a straight segment with the density given as a function on this segment such as this metal rod:



For a numerical function f on segment $[a,b]$ (linear density), we defined the weight as its Riemann integral $\int_a^b f\,dx$.

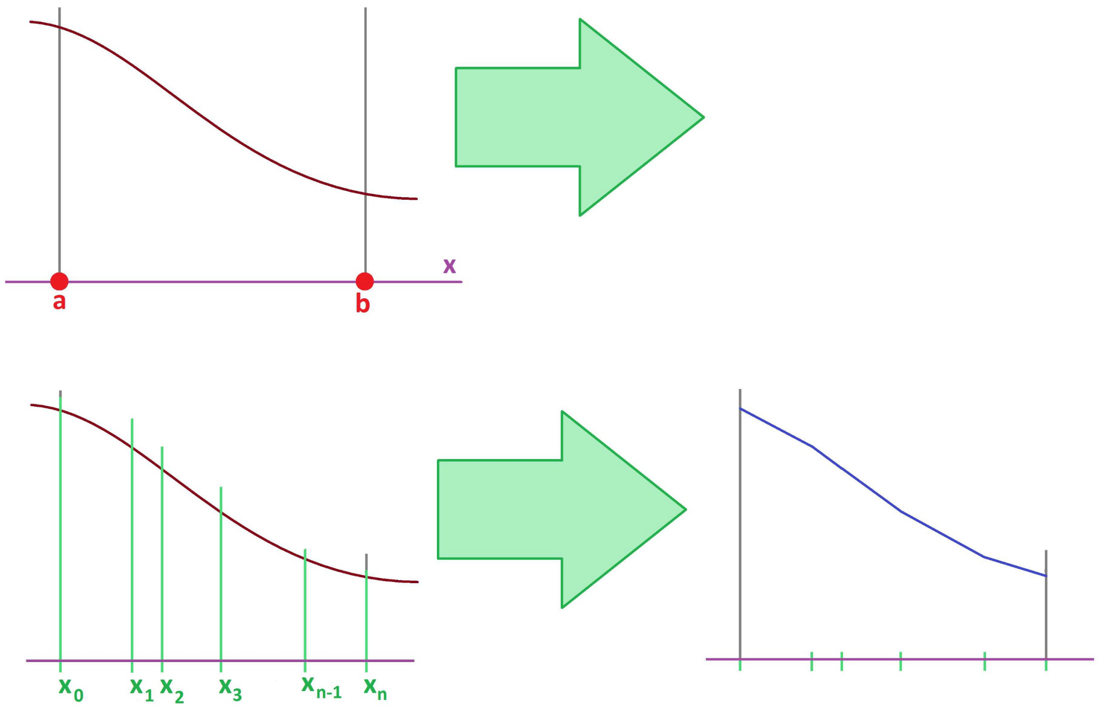
The combination of this analysis and the analysis above provides the solution.

Suppose the curve is used to cut a very thin strip, a wire, from a sheet of metal with variable density. We will define and then compute the *mass of the curve*.

Suppose $X = X(t),\ a \le t \le b$, is a parametric curve in $\mathbf{R}^n$ and the density is given by a function of n variable $z = f(X)$. We partition the interval to sample the curve and then we approximate the pieces of the curve with straight segments and constant densities.

We start with a sampled partition of $[a,b]$ with n intervals: c_k in $[t_{k-1}, t_k],\ k = 1,2,...,n$. With the path of $X = X(t)$ going from $X(t_{k-1})$ to $X(t_k)$, the segment between these two points weighs approximately:

$$f(c_k)||X(t_k) - X(t_{k-1})||\,.$$



Thus, the weight of these segments is:

$$\text{total mass } = \sum_{k=1}^{n} f(c_k)||X(t_k) - X(t_{k-1})||\,.$$

To extract a Riemann sum from this sum, we divide and multiply each term by h, as follows:

$$\begin{aligned}\text{Sum of lengths} \quad &= \sum_{k=1}^{n} f(c_k)||X(t_k) - X(t_{k-1})|| \\ &= \sum_{k=1}^{n} f(c_k)||X(t_k) - X(t_{k-1})|| \cdot \frac{h}{h} \\ &= \sum_{k=1}^{n} f(c_k) \left\| \frac{1}{h}(X(t_k) - X(t_{k-1})) \right\| \cdot h\,.\end{aligned}$$

Definition 2.21.1: weight of curve

The *weight of a curve* given by the path of a regular parametric curve $X = X(t)$ over interval $[a, b]$ with density given by a continuous function $z = f(X)$ is defined to be the following (numerical) integral:

$$\boxed{\int_C f\, ds = \int_a^b f(X(t))||X'(t)||\, dt}$$

if it exists. It is also called an *arc-length integral of f along C*.

Theorem 2.21.2: Weight of Curve

The weight of a curve is defined and defined uniquely if it can be represented (parametrized) as the path of a regular parametric curve.

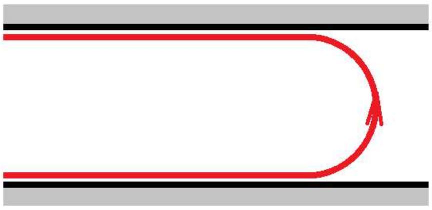
Exercise 2.21.3

Find the weight of the segment of the parabola $y = x^2$ from $(0, 0)$ to $(1, 1)$.

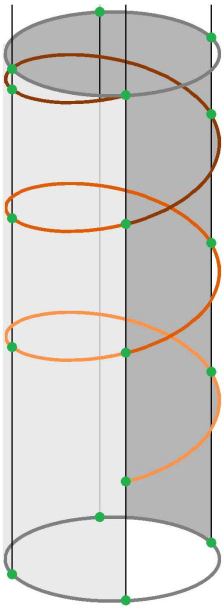
Of course, the integral can also be interpreted as the amount of liquid in a pipe and so on.

2.22. The helix

The turning circle of a car tells us the narrowest width of a street that allows this car to U-turn:



What if instead of a street, we are to build a *circular ramp* in a parking garage?



The shape of such a curve is the *helix*. The standard parametrization of a helix of radius R is given below:

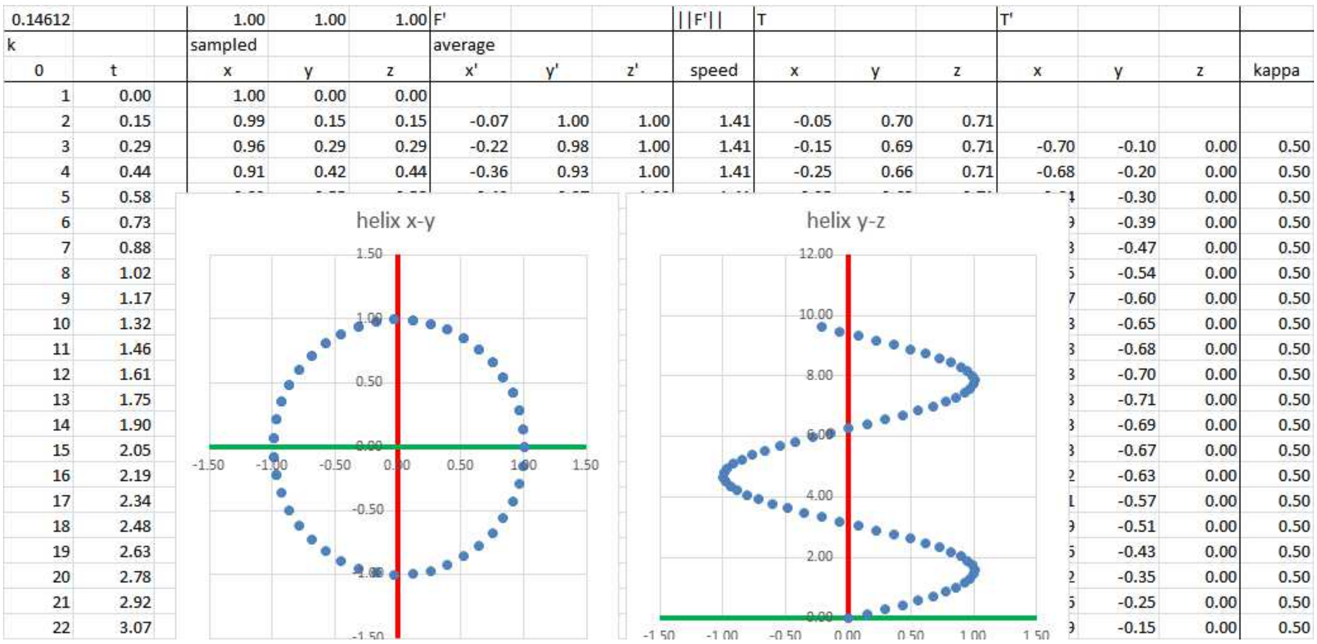
$$F(t) =< R\cos t,\ R\sin t,\ t > \ .$$

It is a combination of a rotation in the horizontal plane and a vertical ascend.

Now, suppose the turning circle of our car has radius r. Also, suppose that our helical ramp is contained in a cylinder of radius R. What should the radius of this cylinder be to accommodate our car? Larger, smaller, or the same?

The answer is *smaller*, because some of the turn is carried out in the vertical direction. The smaller cylinder doesn't have to contain this larger circle because such a circle approximates the curve one point at a time.

We take the spreadsheet for the circle and simply add a z-column wherever we have x- and y-columns. The computed magnitudes – including the curvature itself – are then adjusted to include the third component.



0.14612		1.00	1.00	1.00	F'			||F'||	T			T'			
k		sampled			average										
0	t	x	y	z	x'	y'	z'	speed	x	y	z	x	y	z	kappa
1	0.00	1.00	0.00	0.00											
2	0.15	0.99	0.15	0.15	-0.07	1.00	1.00	1.41	-0.05	0.70	0.71				
3	0.29	0.96	0.29	0.29	-0.22	0.98	1.00	1.41	-0.15	0.69	0.71	-0.70	-0.10	0.00	0.50
4	0.44	0.91	0.42	0.44	-0.36	0.93	1.00	1.41	-0.25	0.66	0.71	-0.68	-0.20	0.00	0.50
5	0.58												-0.30	0.00	0.50
6	0.73												-0.39	0.00	0.50
7	0.88												-0.47	0.00	0.50
8	1.02												-0.54	0.00	0.50
9	1.17												-0.60	0.00	0.50
10	1.32												-0.65	0.00	0.50
11	1.46												-0.68	0.00	0.50
12	1.61												-0.70	0.00	0.50
13	1.75												-0.71	0.00	0.50
14	1.90												-0.69	0.00	0.50
15	2.05												-0.67	0.00	0.50
16	2.19												-0.63	0.00	0.50
17	2.34												-0.57	0.00	0.50
18	2.48												-0.51	0.00	0.50
19	2.63												-0.43	0.00	0.50
20	2.78												-0.35	0.00	0.50
21	2.92												-0.25	0.00	0.50
22	3.07												-0.15	0.00	0.50

The curvature is $\varkappa = 1/2$ and, therefore, the radius of curvature is $R = 2$. So the helix is approximated by a circle of twice its radius!

Let's provide an algebraic result.

Theorem 2.22.1: Curvature of Helix

The curvature of the helix of radius R is constant and equal to:

$$\boxed{\varkappa = \frac{R}{R^2+1}}$$

It follows, first, that

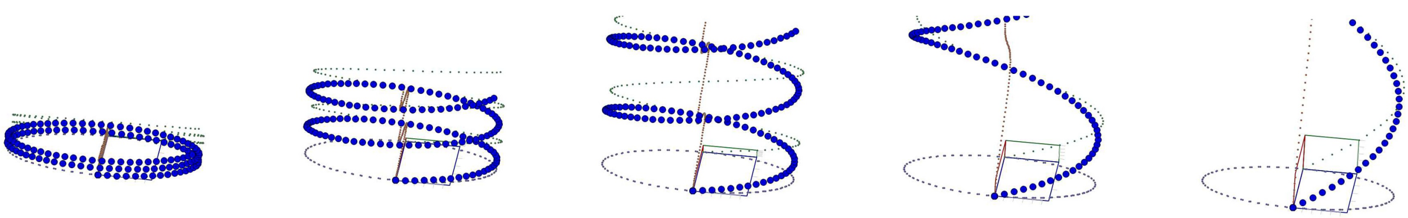
$$\varkappa \to 0 \quad \text{as } R \to \infty\,.$$

This means that a widening helix looks more and more like a circle, larger and larger and, eventually, like a straight line.

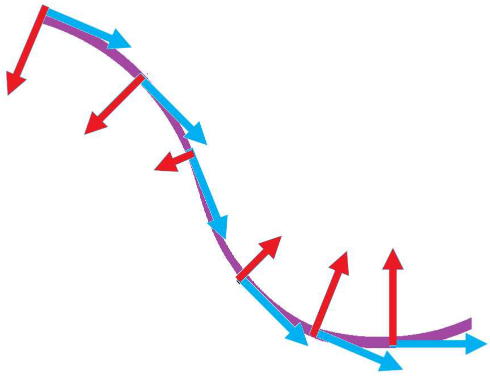
Second, we have:

$$\varkappa \to 0 \quad \text{as } R \to 0\,.$$

This means that a narrowing helix looks more and more like a vertical line. Indeed, this is what happens if we increase the rate of ascend:



So the radius of the circle of curvature is 2 but where is this circle located? On the line perpendicular to the curve, i.e., to the tangent line. Since $||T|| = 1$, we can just choose T'!



Definition 2.22.2: unit normal vector

The *unit normal vector* of a parametric curve $X = F(t)$ is defined to be

$$\boxed{N(t) = \frac{T'(t)}{||T'(t)||}}$$

We now repeat computations of the spreadsheet with our calculus tools. We take the standard helix of radius 1 and differentiate it:

$$F(t) =< \cos t,\ \sin t,\ t > \Longrightarrow\ F'(t) =< -\sin t,\ \cos t,\ 1 > \ .$$

The speed next:

$$||F'(t)|| = \sqrt{(-\sin t)^2 + (\cos t)^2 + 1} = \sqrt{2}\,,$$

and the unit tangent vector:

$$T(t) = \frac{1}{\sqrt{2}} < -\sin t,\ \cos t,\ 1 > \ .$$

We differentiate that:

$$T'(t) = \frac{1}{\sqrt{2}} < -\cos t,\ -\sin t,\ 0 > \ .$$

It's horizontal! Its magnitude is:

$$||T'(t)|| = \frac{1}{\sqrt{2}}\sqrt{(-\cos t)^2 + (-\sin t)^2 + 0^2} = \frac{1}{\sqrt{2}}\,.$$

It's constant! Then the unit normal vector is:

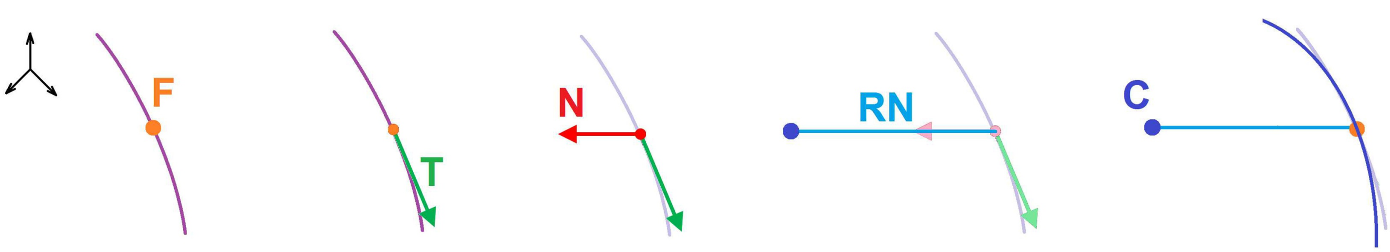
$$N(t) = \frac{1}{\sqrt{2}} < -\cos t,\ -\sin t,\ 0 > \div \frac{1}{\sqrt{2}} =< -\cos t,\ -\sin t,\ 0 > \ .$$

How far do we go? The step is the radius of curvature $R = 2$. The location of the center is:

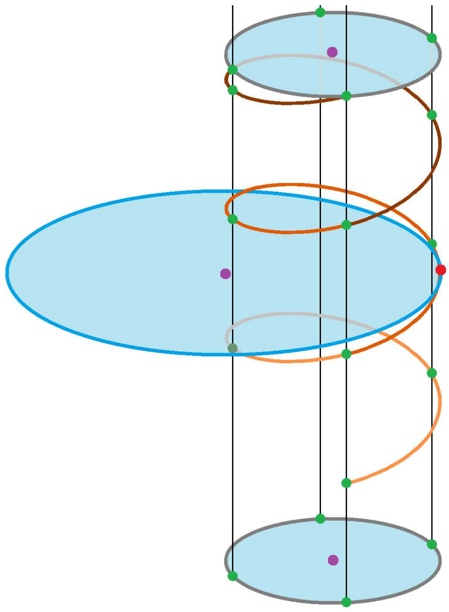
$$C = F(t) + 2N(t) =< \cos t,\ \sin t,\ t > +2 < -\cos t,\ -\sin t,\ 0 >=< -\cos t,\ -\sin t,\ 0 >= -F(t)\,.$$

It's the point exactly opposite to the original location $F(t)$!

For the helix, we are to make a stop from the location, $F(t)$ in this direction, $N(t)$, of length 2.

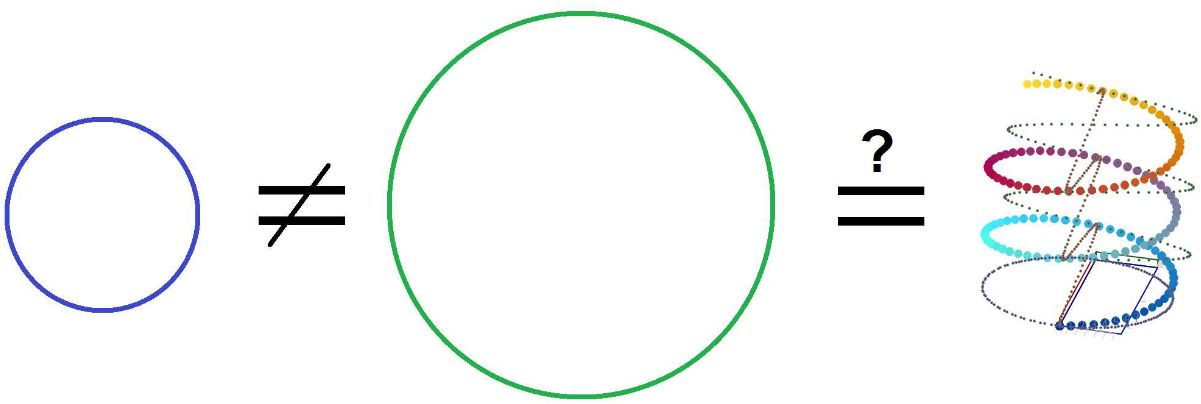


This is the end result: the radius is the double of the radius of the cylinder and, therefore, the center lies on the other side of the cylinder.



Exercise 2.22.3

The curvature of the helix of radius 1 is the same as the circle of radius 2. Can we tell them apart – from the inside?



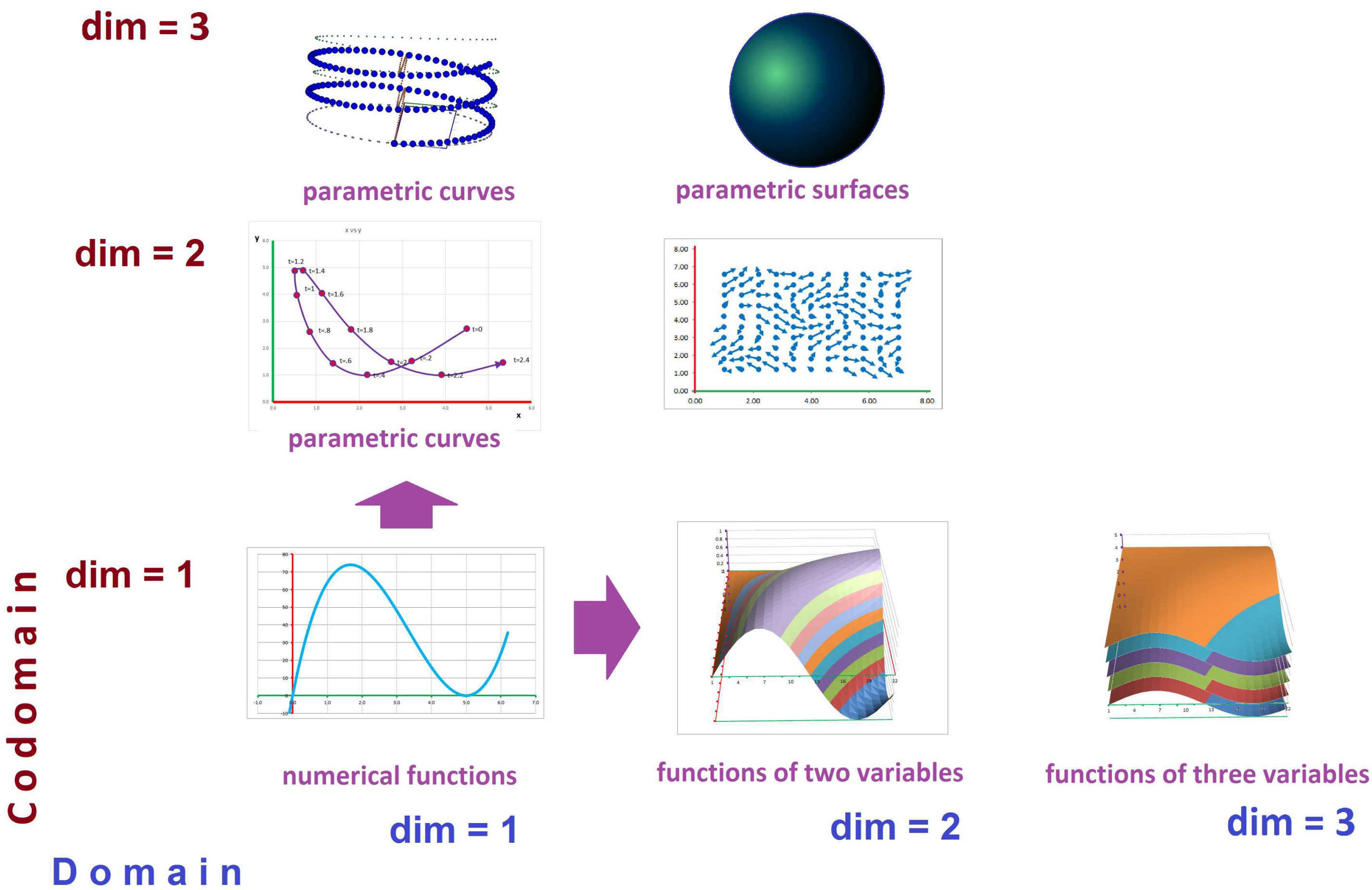
Chapter 3: Functions of several variables

Contents

3.1. Overview of functions

Let's review multidimensional functions.

We have two axes: the dimension of the domain and the dimension of the range:



We covered the very first cell in Volume 2 (Chapter 2DC-3). In Chapter 2, we made a step in the vertical direction and explored the first column of this table.

It is now time to move to the right. We retreat to the first cell because the new material does not depend on the material of Chapter 2 – or vice versa, even though they do interact via compositions. We will not jump diagonally!

We need to appreciate, however, the different challenges these two steps present. Every *two* numerical functions make a planar parametric curve and, conversely, every planar parametric curve is just a pair numerical functions. On the other hand, we can see that the surface that is the graph of a function of two variables produces – through cutting by vertical planes – *infinitely many* graphs of numerical functions.

Note that the first cell has curves but not all of them because some of them fail the vertical line test – such as the circle – and can't be represented by graphs of numerical functions. Hence the need for parametric curves. Similarly, the first cell of the second column has surfaces but not all of them because some of them fail the vertical line test – such as the sphere – and can't be represented by graphs of functions of two variables. Hence the need for parametric surfaces, shown higher in this column. They are presented in Chapter 4.

We represent a function diagrammatically as a *black box* that processes the input and produces the output of whatever nature:

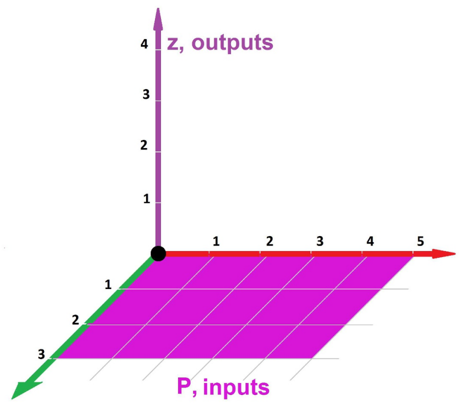
$$\begin{array}{ccccc} \text{input} & & \text{function} & & \text{output} \\ x & \mapsto & \boxed{f} & \mapsto & y \end{array}$$

Let's compare the two:

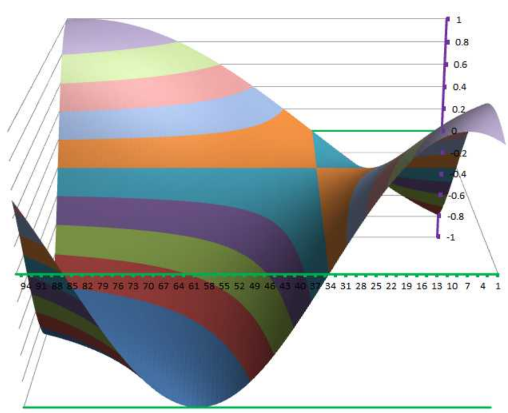
		parametric				function of	
input		curve		output	input	two variables	output
t	$\mapsto$	$\boxed{F}$	$\mapsto$	X	$X \mapsto$	$\boxed{f}$	$\mapsto z$
$\mathbf{R}$				$\mathbf{R}^m$	$\mathbf{R}^m$		$\mathbf{R}$
number				point or vector	point or vector		number

They can be linked up in the middle, producing a composition.

But our interest is the latter starting with dimension $m = 2$:

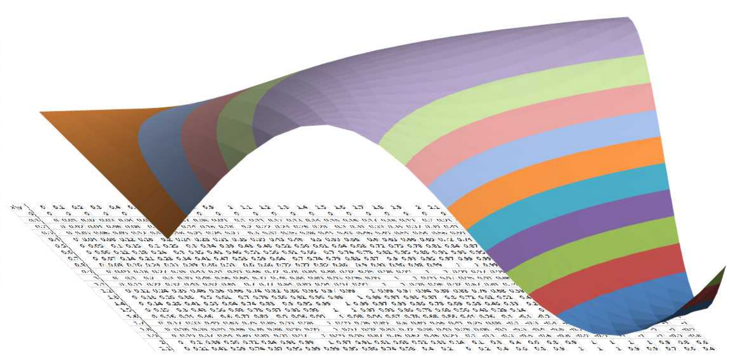


The main metaphor for a function of two variables will remain to be *terrain*:



Here, every pair $X = (x, y)$ represents a location on a map and $z = f(x, y)$ is the elevation of the terrain at that location. This is how it is plotted by a spreadsheet:

x\y	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28
0.2	0	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24	0.26	0.28	0.3	0.31	0.33	0.35	0.37	0.39	0.41	0.43	0.44	0.46	0.48	0.5	0.51	0.53
0.3	0	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.3	0.32	0.35	0.38	0.41	0.43	0.46	0.49	0.51	0.54	0.56	0.59	0.61	0.64	0.66	0.68	0.7	0.72	0.74
0.4	0	0.04	0.08	0.12	0.16	0.2	0.24	0.28	0.31	0.35	0.39	0.43	0.46	0.5	0.53	0.56	0.6	0.63	0.66	0.69	0.72	0.74	0.77	0.8	0.82	0.84	0.86	0.88	0.9
0.5	0	0.05	0.1	0.15	0.2	0.25	0.3	0.34	0.39	0.43	0.48	0.52	0.56	0.61	0.64	0.68	0.72	0.75	0.78	0.81	0.84	0.87	0.89	0.91	0.93	0.95	0.96	0.98	0.99
0.6	0	0.06	0.12	0.18	0.24	0.3	0.35	0.41	0.46	0.51	0.56	0.61	0.66	0.7	0.74	0.78	0.82	0.85	0.88	0.91	0.93	0.95	0.97	0.98	0.99	1	1	1	0.99
0.7	0	0.07	0.14	0.21	0.28	0.34	0.41	0.47	0.53	0.59	0.64	0.7	0.74	0.79	0.83	0.87	0.9	0.93	0.95	0.97	0.99	0.99	1	1	0.99	0.98	0.97	0.95	0.93
0.8	0	0.08	0.16	0.24	0.31	0.39	0.46	0.53	0.6	0.66	0.72	0.77	0.82	0.86	0.9	0.93	0.96	0.98	0.99	1	1	0.99	0.98	0.96	0.94	0.91	0.87	0.83	0.78
0.9	0	0.09	0.18	0.27	0.35	0.43	0.51	0.59	0.66	0.72	0.78	0.84	0.88	0.92	0.95	0.98	0.99	1	1	0.99	0.97	0.95	0.92	0.88	0.83	0.78	0.72	0.65	0.58
1	0	0.1	0.2	0.3	0.39	0.48	0.56	0.64	0.72	0.78	0.84	0.89	0.93	0.96	0.99	1	1	0.99	0.97	0.95	0.91	0.86	0.81	0.75	0.68	0.6	0.52	0.43	0.33
1.1	0	0.11	0.22	0.32	0.43	0.52	0.61	0.7	0.77	0.84	0.89	0.94	0.97	0.99	1	1	0.98	0.96	0.92	0.87	0.81	0.74	0.66	0.57	0.48	0.38	0.28	0.17	0.06
1.2	0	0.12	0.24	0.35	0.46	0.56	0.66	0.74	0.82	0.88	0.93	0.97	0.99	1	0.99	0.97	0.94	0.89	0.83	0.76	0.68	0.58	0.48	0.37	0.26	0.14	0.02	-0.1	-0.2
1.3	0	0.13	0.26	0.38	0.5	0.61	0.7	0.79	0.86	0.92	0.96	0.99	1	0.99	0.97	0.93	0.87	0.8	0.72	0.62	0.52	0.4	0.28	0.15	0.02	-0.1	-0.2	-0.4	-0.5
1.4	0	0.14	0.28	0.41	0.53	0.64	0.74	0.83	0.9	0.95	0.99	1	0.99	0.97	0.93	0.86	0.78	0.69	0.58	0.46	0.33	0.2	0.06	-0.1	-0.2	-0.4	-0.5	-0.6	-0.7
1.5	0	0.15	0.3	0.43	0.56	0.68	0.78	0.87	0.93	0.98	1	1	0.97	0.93	0.86	0.78	0.68	0.56	0.43	0.29	0.14	-0	-0.2	-0.3	-0.4	-0.6	-0.7	-0.8	-0.9
1.6	0	0.16	0.31	0.46	0.6	0.72	0.82	0.9	0.96	0.99	1	0.98	0.94	0.87	0.78	0.68	0.55	0.41	0.26	0.1	-0.1	-0.2	-0.4	-0.5	-0.6	-0.8	-0.9	-0.9	-1
1.7	0	0.17	0.33	0.49	0.63	0.75	0.85	0.93	0.98	1	0.99	0.96	0.89	0.8	0.69	0.56	0.41	0.25	0.08	-0.1	-0.3	-0.4	-0.6	-0.7	-0.8	-0.9	-1	-1	-1
1.8	0	0.18	0.35	0.51	0.66	0.78	0.88	0.95	0.99	1	0.97	0.92	0.83	0.72	0.58	0.43	0.26	0.08	-0.1	-0.3	-0.4	-0.6	-0.7	-0.8	-0.9	-1	-1	-1	-0.9
1.9	0	0.19	0.37	0.54	0.69	0.81	0.91	0.97	1	0.99	0.95	0.87	0.76	0.62	0.46	0.29	0.1	-0.1	-0.3	-0.5	-0.6	-0.8	-0.9	-0.9	-1	-1	-1	-0.9	-0.8
2	0	0.2	0.39	0.56	0.72	0.84	0.93	0.99	1	0.97	0.91	0.81	0.68	0.52	0.33	0.14	-0.1	-0.3	-0.4	-0.6	-0.8	-0.9	-1	-1	-1	-1	-0.9	-0.8	-0.6
2.1	0	0.21	0.41	0.59	0.74	0.87	0.95	0.99	0.99	0.95	0.86	0.74	0.58	0.4	0.2	-0	-0.2	-0.4	-0.6	-0.8	-0.9	-1	-1	-1	-0.9	-0.9	-0.7	-0.6	-0.4



Now *calculus.*

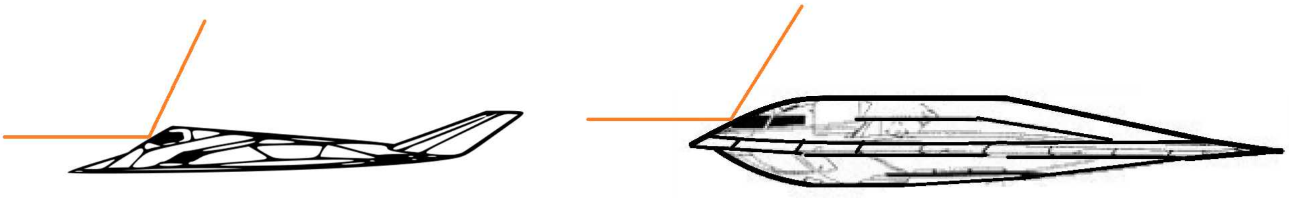
One subject of calculus is change and *motion* and, among others, we will address the question: if a drop of water land on this surface, in what direction will it flow? We will also consider the issue of the rate of change of the function – in any direction.

Secondly, calculus studies tangency and *linear approximations* and, among others, we will address the question: if we zoom in on a particular location on the surface, what does it look like? The short answer is: like a plane. It is discussed in the next section.

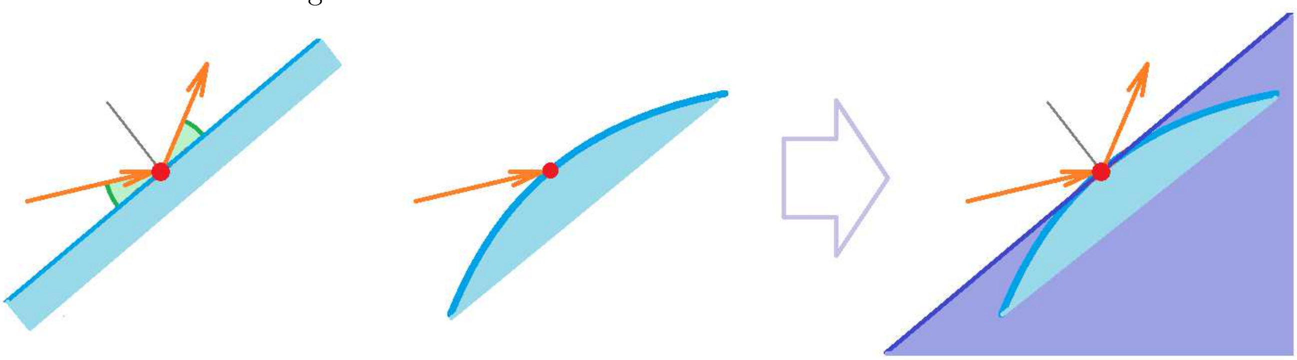
Examples of this issue have been seen previously. Indeed, recall that the *Tangent Problem* asks for a tangent line to a curve at a given point. It has been solved for parametric curves in Chapter 2. However, in real life we see *surfaces* rather than curves. The examples are familiar.

Example 3.1.1: radar

In which direction a radar signal will bounce off a plane when the surface of the plane is curved?



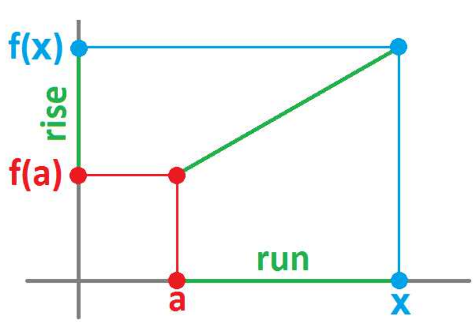
In what direction will light bounce off a curved mirror?



What if it is a whole building?

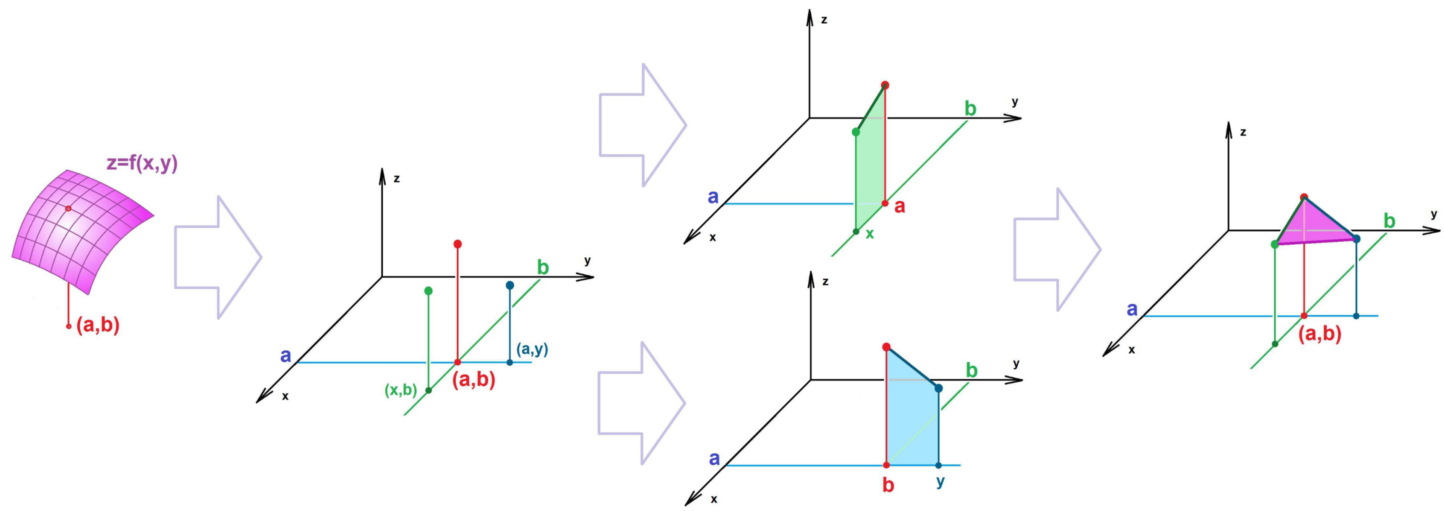
Recall the 1-dimensional case. The difference quotient of a function $y = f(x)$ at $x = a$ is defined as the slope of the line that connects $(a, f(a))$ to the next point $(x, f(x))$:

$$\frac{\Delta f}{\Delta x} = \frac{f(x) - f(a)}{x - a}\,.$$



Now, let's see how this plan applies to functions of two variables $z = f(x, y)$.

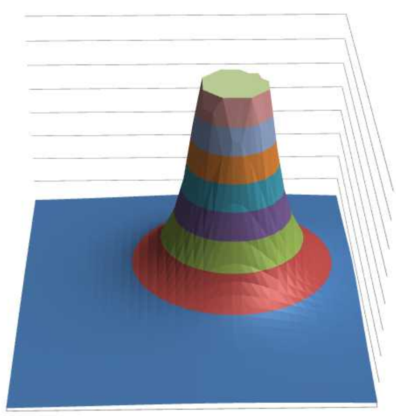
If we are interested in point $(a, b, f(a, b))$ on the graph of f, we still plot the line that connects this point to the next point on the grid.
There are *two* point this time; they lie in the x- and the y-directions from (a, b), i.e., (x, b) and (a, y) with $x \neq a$ and $y \neq b$.



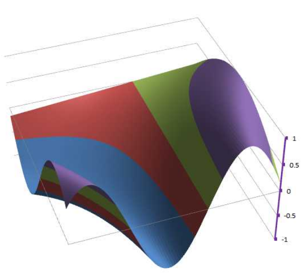
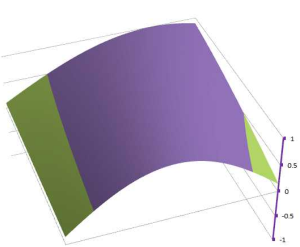
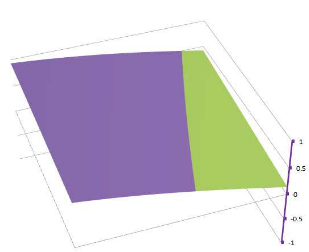
The two slopes in these two directions are the two difference quotients, with respect to x and with respect to y:

$$\frac{\Delta f}{\Delta x} = \frac{f(x, b) - f(a, b)}{x - a} \quad \text{and} \quad \frac{\Delta f}{\Delta y} = \frac{f(a, y) - f(a, b)}{y - b}\,.$$

When done with every pair of nodes on the graph, the result is a *mesh* of triangles:




Furthermore, if the surface is the graph of a continuous function and we zoom in closer and closer on a particular point, we might expect the surface to start to look more and more straight like a *plane*.

3.2. Linear functions: lines in $\mathbf{R}^2$ and planes in $\mathbf{R}^3$

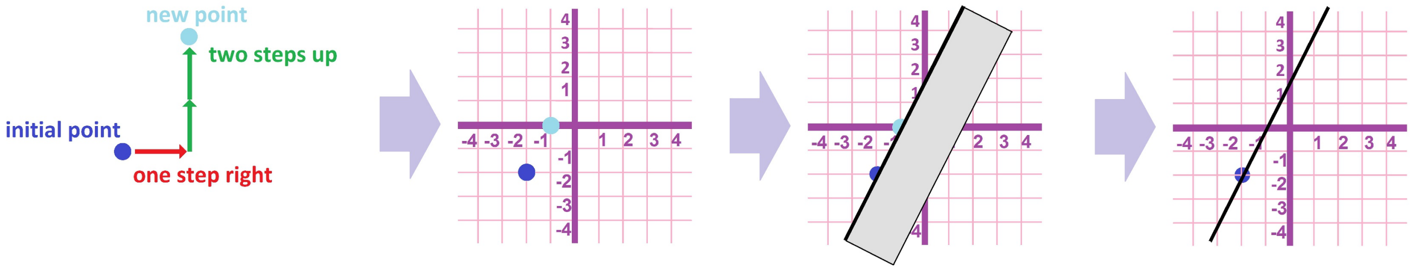
The standard, slope-intercept, form of the equation of a line in the xy-plane is:

$$y = mx + p\,.$$

Here m is the slope and p the y-intercept. Next, the point-slope form of the equation of a line in the xy-plane is:

$$y - b = m(x - a)\,.$$

This is how we can plot this line. We start with the point (a, b) in $\mathbf{R}^2$. Then we make a step along the x-axis with the slope m, i.e., we end up at $(a+1, b+m)$ or $(a+1/m, b+1)$, etc. These two points determine the line.



We also need the general (implicit) equation of a line in $\mathbf{R}^2$:

$$m(x - a) + n(y - b) = 0\,.$$

Let's take a careful look at the left-hand side of this expression. There is a symmetry here over the three coordinates:

$$\begin{array}{ccl} m & \cdot & (x-a)+ \\ n & \cdot & (y-b) = 0 \end{array} \quad \Longleftrightarrow \quad \begin{bmatrix} m \\ n \end{bmatrix} \cdot \begin{bmatrix} x-a \\ y-b \end{bmatrix} = 0\,.$$

This is the *dot product*! One can think of the second vector as the *increment* of the independent, vector, variable:

$$\begin{bmatrix} x-a \\ y-b \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix}.$$

The equation becomes:

$$< m, n > \cdot \Big((x, y) - (a, b) \Big) = 0\,.$$

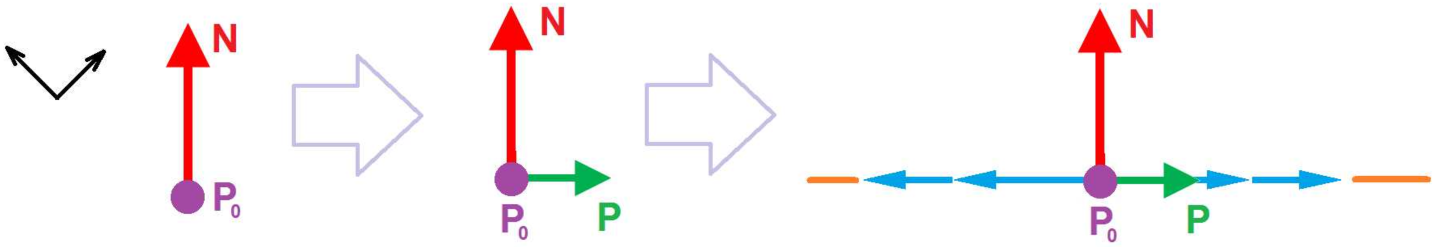
Finally a coordinate-free version:

$$\boxed{N \cdot (P - P_0) = 0 \quad \text{or} \quad N \cdot P_0P = 0}$$

Here we have in $\mathbf{R}^2$:

- P is the variable point.
- P_0 is the fixed point.
- N is any vector that represents the slope of the line.

The meaning of the vector N is revealed once we remember that the dot product of two vectors is 0 if and only if they are perpendicular.



Definition 3.2.1: line through point with normal vector

Suppose a point P_0 and a non-zero vector N are given on the plane. Then the *line through* P_0 *with normal vector* N is the collection of all points P – and P_0 itself – that satisfy:

$$P_0P \perp N\,.$$

Now the *planes* in space. We will approach the issue in a manner analogous to that for lines.

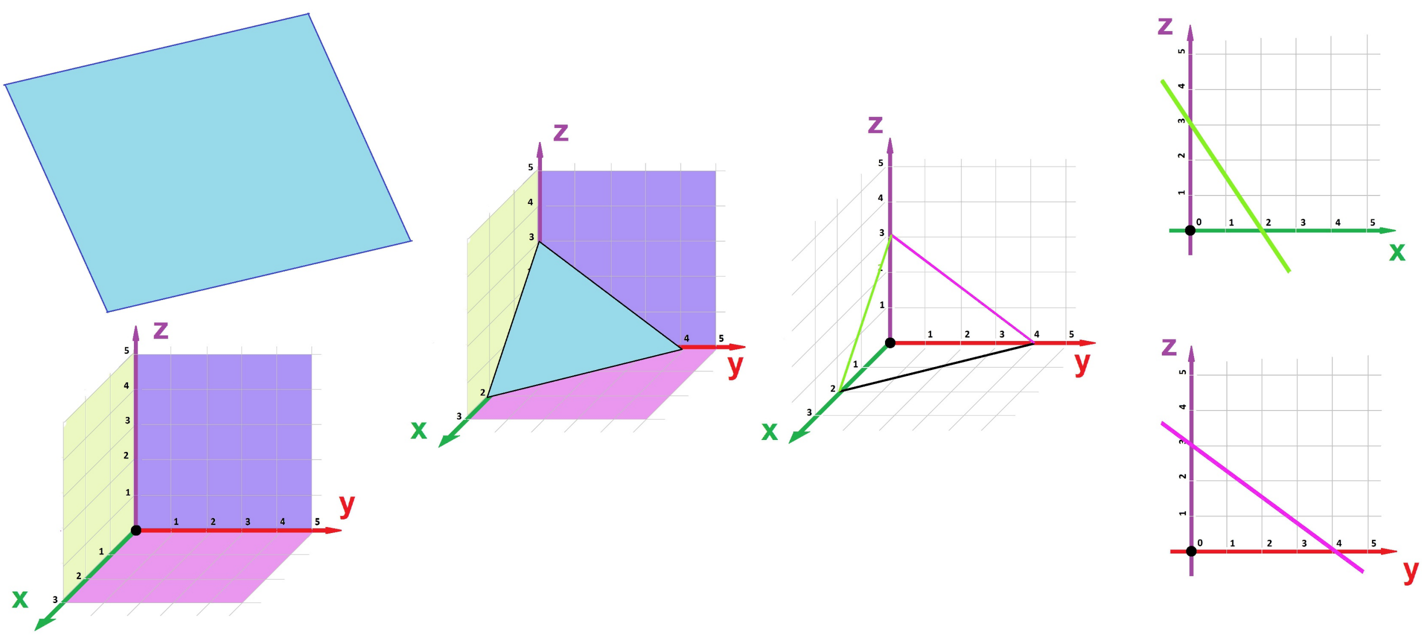
The slope-intercept form of the equation of a line in the xy-plane:

$$y = mx + p\,.$$

has an analogue. A similar, also in some sense *slope-intercept*, form of the equation of a plane in $\mathbf{R}^3$ is:

$$z = mx + ny + p\,.$$

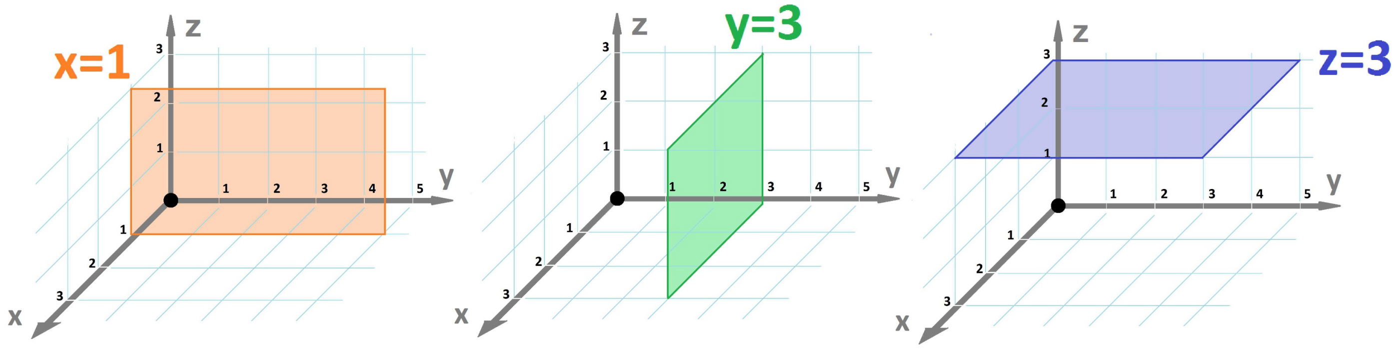
Indeed, if we substitute $x = y = 0$ we have $z = p$. Then p is the z-intercept!



The “slopes” depicted are $n = -3/2$ and $m = -3/4$.

In what sense are m and n the slopes? Let's substitute $y = 0$ first. We have $z = mx + p$, an equation of a line – in the xz-plane. Its slope is m. Now we substitute $x = 0$. We have $z = ny + p$, an equation of a line – in the yz-plane. Its slope is n. Now, if we cut the plane with any plane parallel to the xz-plane, or respectively yz-plane, the resulting line has the same slope; for example:

$$y = 1 \implies z = mx + n + p; \quad x = 1 \implies z = m + ny + p\,.$$



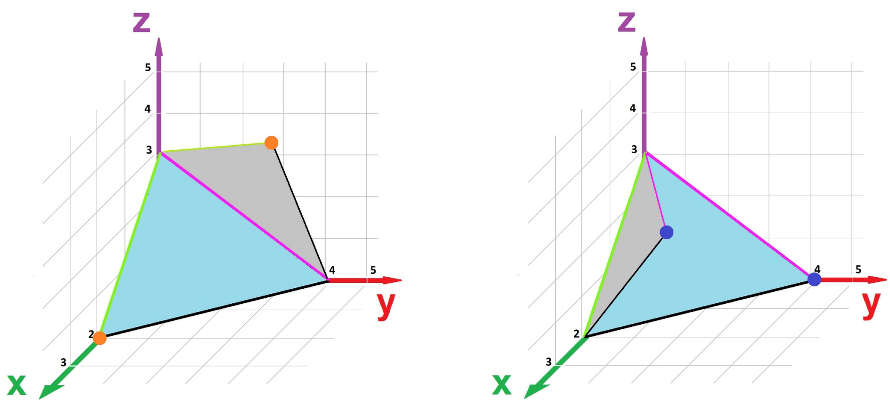
Therefore, we are justified to say that m and n are the *slopes* of the plane with respect to the variables x and y respectively:

$$z = \underbrace{m}_{x\text{-slope}} \cdot x + \underbrace{n}_{y\text{-slope}} \cdot y + \underbrace{p}_{z\text{-intercept}}$$

Cutting with a horizontal plane also produces a line:

$$z = 1 \implies 1 = mx + ny + p\,.$$

These two slopes are also independent! We can use the line of the intersection of the plane with one of the coordinate planes as a hinge – so that *this* slope remains the same – and rotate the plane as a door so that the *other* slope will change:



The analogue of the point-slope form of the equation of a line in the xy-plane:

$$y - b = m(x - a)\,.$$

is as follows. This is a similar, also in some sense *point-slope*, form of the equation of a plane. This is the step we make:

in $\mathbf{R}^2$	in $\mathbf{R}^3$
point (a, b)	point (a, b, c)
slope m	slopes m, n

This analogy produces the following formula:

$$z - c = m(x - a) + n(y - b)\,.$$

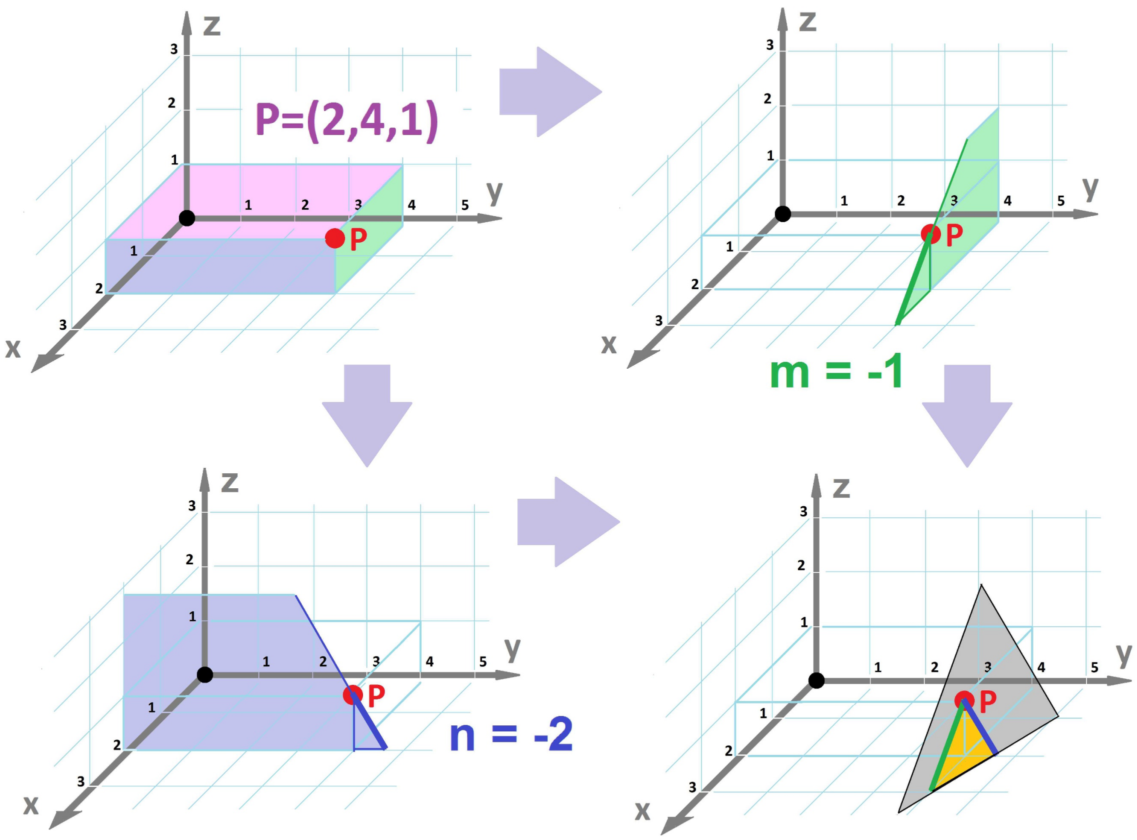
Expanding it takes us back to the point-slope formula.

Example 3.2.2: plane

Let's plot such a plane with:

$$(a, b, c) = (2, 4, 1),\ m = -1,\ n = -2\,.$$

First, we fix $y = 4$ and change x: from the point $(2, 4, 1)$, we make a step along the x-axis with the slope -1, i.e., we end up at $(2 + 1, 4, 1 - 1) = (3, 4, 9)$. We plot this line:



Second, we fix $x = 2$ and change y: from the point $(2, 4, 1)$, we make a step along the y-axis with the slope -2, i.e., we end up at $(2, 4 + 1, 1 - 2) = (2, 5, -1)$. We plot this line. And, finally, we plot a plane through those two lines.

Generally, this is how we plot a plane given by such an equation:

$$z - c = m(x - a) + n(y - b)\,.$$

We start by plotting the point (a, b, c) in $\mathbf{R}^3$. Now we treat one variable at a time. First, we fix y and change x. From the point (a, b, c), we make a step along the x-axis with the x-slope, i.e., we end up at $(a + 1, b, c + m)$ or $(a + 1/m, b, c + 1)$, etc. The equation of this line is:

$$z - c = m(x - a),\ \ y = b\,.$$

Second, we fix x and change y. We make a step along the y-axis with the y-slope, i.e., we end up at $(a, b + 1, c + n)$ or $(a, b + 1/n, c + 1)$, etc. The equation of this line is:

$$x = a,\ \ z - c = n(y - b)\,.$$

These three points (or those two lines through the same point) determine the plane.

What is a plane anyway? The planes we have considered so far are the graphs of (linear) functions of two variables:

$$z = f(x, y) = mx + ny + p\,.$$

They have to satisfy the *Vertical Line Test*. Therefore, the vertical planes – even the two xz- and yz-planes – are excluded. This is very similar to the situation with lines and the impossibility to represent *all* lines in the standard form $y = mx + p$ and we need the general (implicit) equation of a line in $\mathbf{R}^2$:

$$m(x - a) + n(y - b) = 0\,.$$

The general (implicit) equation of a line in $\mathbf{R}^3$ is:

$$m(x - a) + n(y - b) + k(z - c) = 0\,.$$

Let's take a careful look at the equation:

$$\begin{array}{lll} m & \cdot & (x-a)+ \\ n & \cdot & (y-b)+ \\ k & \cdot & (z-c)=0 \end{array} \qquad \Longleftrightarrow \qquad \begin{bmatrix} m \\ n \\ k \end{bmatrix} \cdot \begin{bmatrix} x-a \\ y-b \\ z-c \end{bmatrix} = 0 \,.$$

This is the *dot product*! The equation becomes:

$$< m, n, k > \cdot < x-a, y-b, z-c >= 0 \,,$$

or even better:

$$< m, n, k > \cdot \Big((x, y, z) - (a, b, c) \Big) = 0 \,.$$

One can think of the last vector as the *increment* of the independent variable (x, y, z).

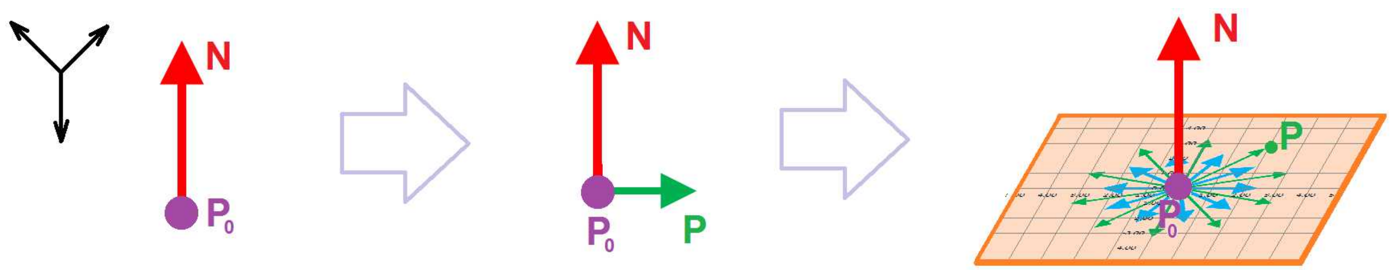
Finally, we have a coordinate-free version of our equation:

$$\boxed{N \cdot (P - P_0) = 0 \quad \text{or} \quad N \cdot P_0P = 0}$$

It is exactly the same as the one for the line! Here we have in $\mathbf{R}^3$:

- P is the variable point.
- P_0 is the fixed point.
- N is any vector that somehow represents the slope of the plane.

The construction is outlined below:



These vectors are like bicycle's *spokes* to the hub N. This idea gives us our definition.

Definition 3.2.3: plane through point with normal vector

Suppose a point P_0 and a non-zero vector N are given. Then the *plane through* P_0 *with normal vector* N is the collection of all points P – and P_0 itself – that satisfy:

$$P_0P \perp N \,.$$

This definition gives us different results in different dimensions:

dimension	ambient space	"hyperplane"	
2	$\mathbf{R}^2$	$\mathbf{R}^1$	line
3	$\mathbf{R}^3$	$\mathbf{R}^2$	plane
4	$\mathbf{R}^4$	$\mathbf{R}^3$	–
...	...	...	...

A hyperplane is something very "thin" relative the whole space but not as thin as, say, a curve.

This wide applicability shows that learning the dot product really pays off!

We will need this formula to study parametric surfaces later. In this chapter, we limit ourselves to functions of several variables and, therefore, non-vertical planes. What makes a plane non-vertical? A non-zero vertical component of the normal vector. Since length don't matter here (only the angles), we can simply assume that this component is equal to one:

$$N =< m, n, 1 > \ .$$

Then the equation of a plane simplifies:

$$0 = N \cdot (P - P_0) =< m, n, 1 > \cdot < x-a, y-b, z-c >= m(x-a) + n(y-b) + z - c\,,$$

or the familiar

$$z = c + m(x-a) + n(y-b)\,.$$

In the vector notation, we have:

$$z = c + M \cdot (Q - Q_0)\,.$$

In case of dimension 2 we have here:

- $Q = (x, y)$ is the variable point.
- $Q_0 = (a, b)$ is the fixed point.
- $M =< m, n >$ is the vector the components of which are the two slopes of the plane.

This is a *linear function*:

$$z = f(x, y) = c + m(x-a) + n(y-b) = p + mx + ny\,,$$

and $M =< m, n >$ is called the *gradient* of f.

3.3. An example of a non-linear function

What makes a function *linear*? The implicit answer has been: the variable can only be multiplied by a constant and added to a constant. The first part of the answer still applies, even to functions of many variables as it prohibits multiplication by another variable. You can still add them.

The non-linearity of a function of two variables is seen as soon as it is plotted – it's not a plane – but it also suffices to *limit its domain*.

So, this function isn't linear:

$$f(x, y) = xy\,.$$

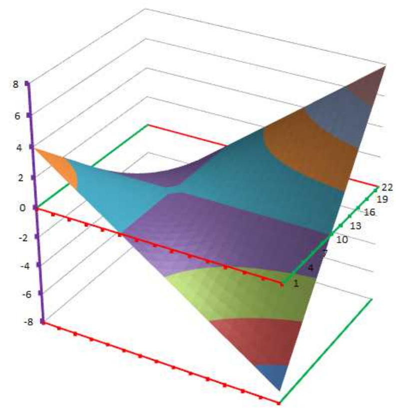
In fact, $xy = x^1y^1$ is seen as *quadratic* if we add the powers of x and y: $1 + 1 = 2$. We come to the same conclusion when we limit the domain of the function to the line $y = x$ in the xy-plane; using $y = x$ s a substitution we arrive to a function of one variable:

$$g(x) = f(x, x) = x \cdot x = x^2\,.$$

So, the part of the graph of f that lies exactly above the line $y = x$ is a *parabola*. And so is the part that lies above the line $y = -x$; it's just open down instead of up:

$$h(x) = f(x, -x) = x \cdot (-x) = -x^2\,.$$

These two parabolas have a single point in common and therefore make up the essential part of a *saddle point*:

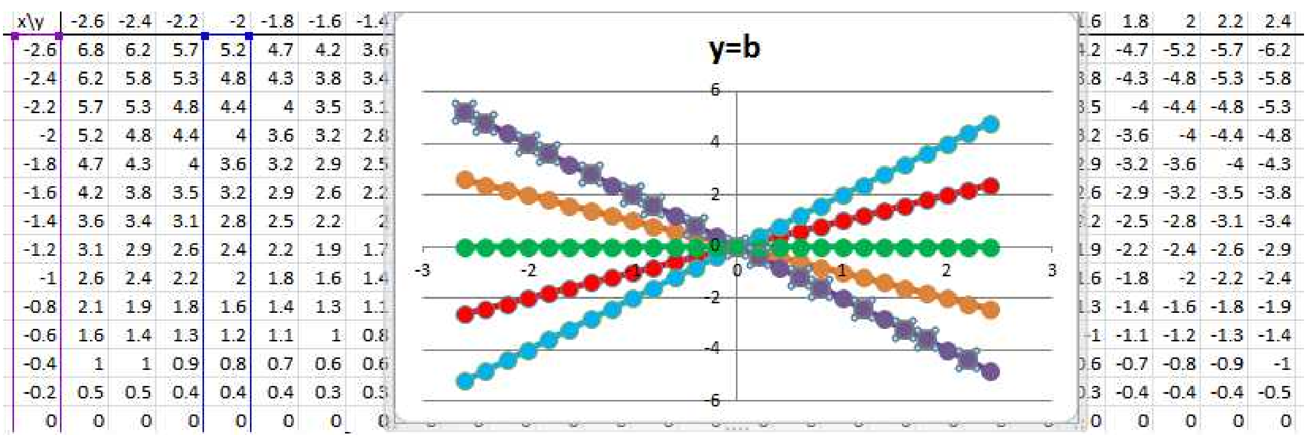


The former parabola gives room for the horse's front and back while the latter for the horseman's legs.

A simpler way to limit the domain is to fix one independent variable at a time. We fix y first:

plane	equation	curve
$y=2$	$z=x\cdot 2$	line with slope 2
$y=1$	$z=x\cdot 1$	line with slope 1
$y=0$	$z=x\cdot 0=0$	line with slope 0
$y=-1$	$z=x\cdot(-1)$	line with slope 1
$y=-2$	$z=x\cdot(-2)$	line with slope -2

The view shown below is from the direction of the y-axis:

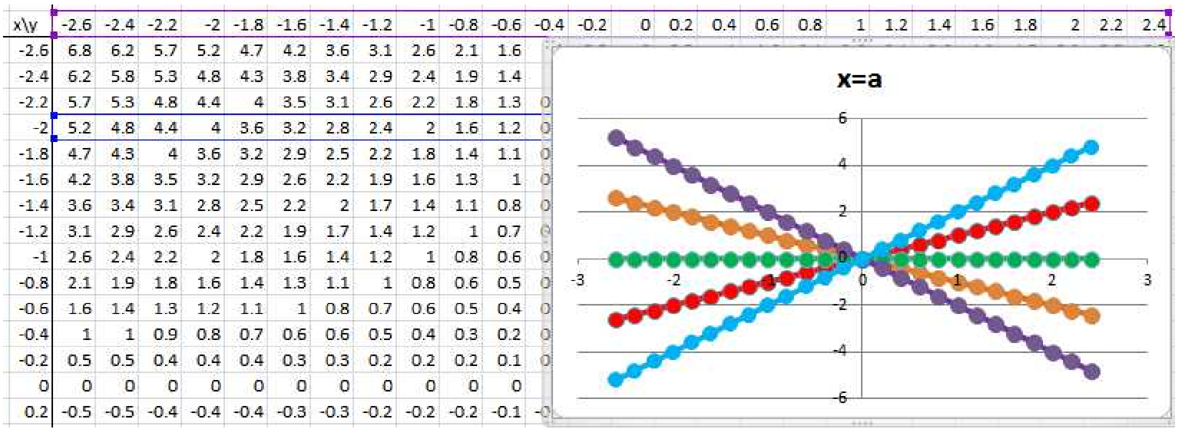


The data for each line comes from the x-column of the spreadsheet and one of the z-columns. These lines give the lines of elevation of this terrain in a particular, say, east-west direction. This is equivalent to cutting the graph by a vertical plane parallel to the xz-plane.

We fix x second:

plane	equation	curve
$x=2$	$z=2\cdot y$	line with slope 2
$x=1$	$z=1\cdot y$	line with slope 1
$x=0$	$z=0\cdot y=0$	line with slope 0
$x=-1$	$z=(-1)\cdot y$	line with slope 1
$x=-2$	$z=(-2)\cdot y$	line with slope -2

This is equivalent to cutting the graph by a vertical plane parallel to the yz-plane. The view shown below is from the direction of the x-axis:



x\y	-2.6	-2.4	-2.2	-2	-1.8	-1.6	-1.4	-1.2	-1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4
-2.6	6.8	6.2	5.7	5.2	4.7	4.2	3.6	3.1	2.6	2.1	1.6															
-2.4	6.2	5.8	5.3	4.8	4.3	3.8	3.4	2.9	2.4	1.9	1.4															
-2.2	5.7	5.3	4.8	4.4	4	3.5	3.1	2.6	2.2	1.8	1.3															
-2	5.2	4.8	4.4	4	3.6	3.2	2.8	2.4	2	1.6	1.2															
-1.8	4.7	4.3	4	3.6	3.2	2.9	2.5	2.2	1.8	1.4	1.1															
-1.6	4.2	3.8	3.5	3.2	2.9	2.6	2.2	1.9	1.6	1.3	1															
-1.4	3.6	3.4	3.1	2.8	2.5	2.2	2	1.7	1.4	1.1	0.8															
-1.2	3.1	2.9	2.6	2.4	2.2	1.9	1.7	1.4	1.2	1	0.7															
-1	2.6	2.4	2.2	2	1.8	1.6	1.4	1.2	1	0.8	0.6															
-0.8	2.1	1.9	1.8	1.6	1.4	1.3	1.1	1	0.8	0.6	0.5															
-0.6	1.6	1.4	1.3	1.2	1.1	1	0.8	0.7	0.6	0.5	0.4															
-0.4	1	1	0.9	0.8	0.7	0.6	0.6	0.5	0.4	0.3	0.2															
-0.2	0.5	0.5	0.4	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.1															
0	0	0	0	0	0	0	0	0	0	0	0															
0.2	-0.5	-0.5	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.2	-0.1															

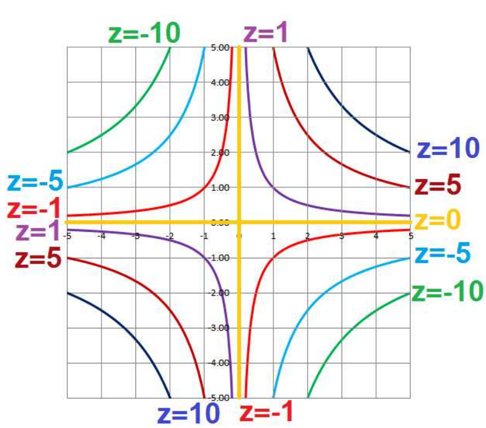
The data for each line comes from the y-row of the spreadsheet and one of the z-rows. These lines give the lines of elevation of this terrain in a particular, say, north-south direction.

Thus the surface of this graph is made of these straight lines! It's a potato chip:

Another way to analyze the graph is to limit the *range* instead of the domain. We fix the dependent variable this time:

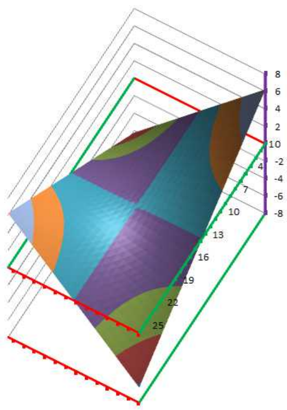
elevation	equation	curve
$z = 2$	$2 = x \cdot y$	hyperbola
$z = 1$	$1 = x \cdot y$	hyperbola
$z = 0$	$0 = x \cdot y$	the two axes
$z = -1$	$-1 = x \cdot y$	hyperbola
$z = -2$	$-2 = x \cdot y$	hyperbola

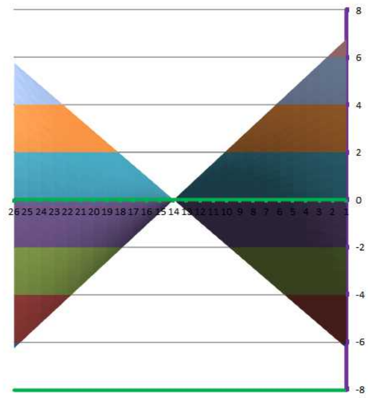
The result is a family of curves:



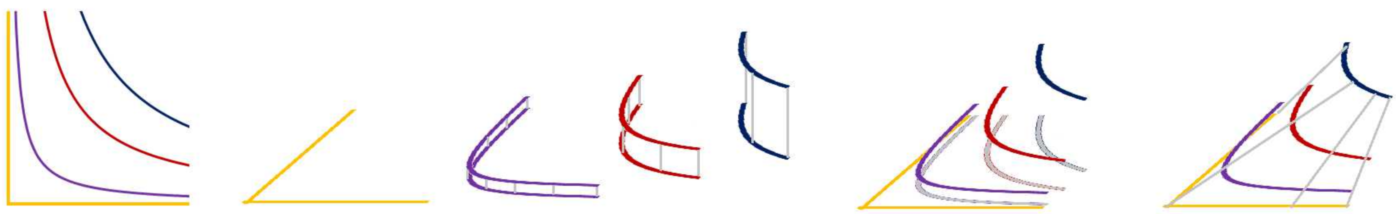
Each is labelled with the corresponding value of z, two branches for each. These lines are the *lines of equal elevation* of this terrain.

We can see who these lines come from cutting the graph by a horizontal plane (parallel to the xy-plane).





We can use them to reassemble the surface by lifting each to the elevation indicated by its label:



In the meantime, the colored parts of the graph correspond to *intervals* of outputs.

This surface presented here is called the *hyperbolic paraboloid*.

3.4. Graphs

Recall the very basic, but very general, definitions.

Definition 3.4.1: graph of function of one variable

The *graph* of a function of one variable $z = f(x)$ is the set of all points in $\mathbf{R}^2$ of the form $(x, f(x))$.

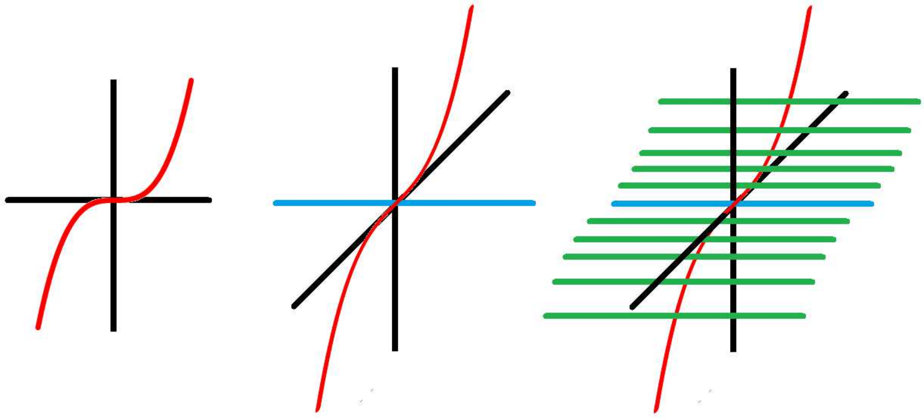
In spite of a few exceptions, the graphs of the function of one variable have been *curves*.

Definition 3.4.2: graph of function

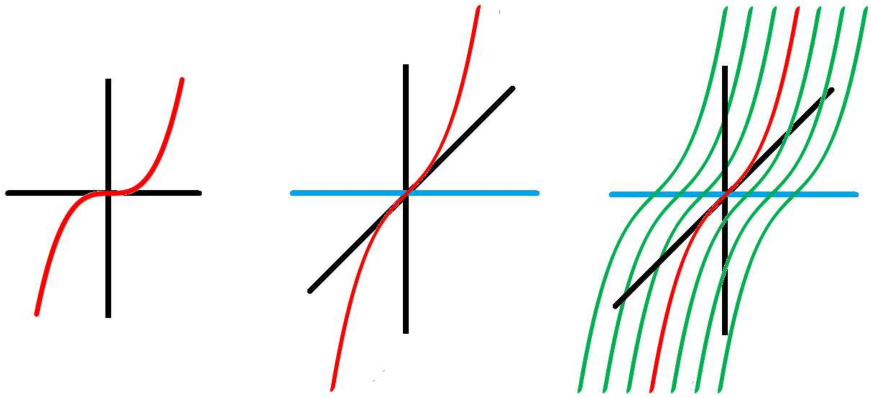
The *graph* of a function of two variables $z = f(x, y)$ is the set of all points in $\mathbf{R}^3$ of the form $(x, y, f(x, y))$.

In spite of possible exceptions, the graphs of functions of two variables we encounter will probably be *surfaces*.

It is important to remember that the theory of functions of several variables include that of the functions of one variable! The formula for f, for example, might have no mention of y such as for example $z = f(x, y) = x^3$. The graph of such a function will be feature-less, i.e., constant, in the y-direction. It will look as if made of planks like a park bench:



In fact, the graph can be acquired from the graph of $z = x^2$ (the curve) by shifting it in the y-direction producing the surface:



In the last section we fixed one independent variable at a time making a function of *two* variables a function of *one* variable subject to the familiar methods. The idea applies to higher dimensions.

Definition 3.4.3: variable function

Suppose $z = f(x_1, ..., x_n)$ is a function of several variables, i.e., $x_1, ..., x_n$ and z are real numbers. Then for each value of $k = 1, 2, ..., n$ and any collection of numbers $x_i = a_i,\ i \neq k$, the numerical function defined by:

$$z = h(x) = f(a_1, ..., a_{k-1}, x, a_{k+1}, ..., a_n)\,,$$

is called the kth *variable function* of f. When $n = 2$, its graph is called the kth *variable curve*.

We thus fix all variables but one making the function of n variables a function of a single variable.

Exercise 3.4.4

What happens when $n = 1$?

Meanwhile, fixing the value of the dependent variable doesn't have to produce a new function. The result is instead an *implicit relation*. The idea applies to higher dimensions too.

Definition 3.4.5: level set

Suppose $z = f(X)$ is a function of several variables, i.e., X belongs to some $\mathbf{R}^n$ and z is a real number. Then for each value of $z = c$, the subset

$$\{X :\ f(X) = c\}$$

of $\mathbf{R}^n$ is called a *level set* of f.

When $n = 2$, it is informally called a *level curve* or a *contour curve*.

Exercise 3.4.6

What happens when $n = 1$?

In general, a level set doesn't have to be a curve as the example of $f(x, y) = 1$ shows.

Theorem 3.4.7: Level Sets

Level sets don't intersect.

Proof.

It follows from the *Vertical Line Test*.

Exercise 3.4.8

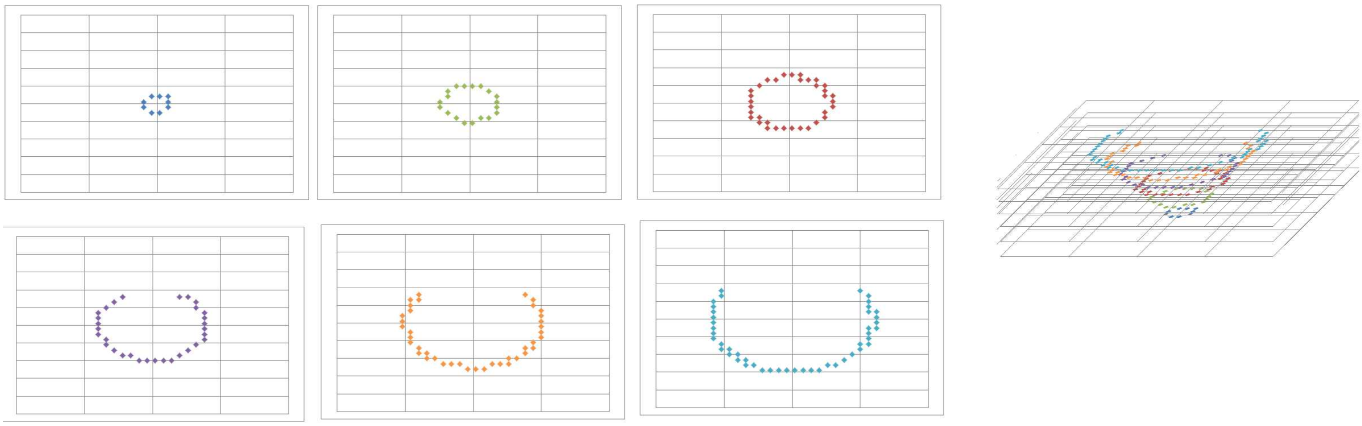
Provide the proof.

Example 3.4.9: level curves

Let's consider:

$$f(x, y) = \sqrt{x^2 + y^2}\,.$$

The level curves are implicit curves and they are plotted point by point:



...unless we recognize them:

$$c = \sqrt{x^2 + y^2} \implies c^2 = x^2 + y^2\,.$$

They are circles! This surface is a half of a *cone*. How do we know it's a cone and not another shape? We limit the domain to this line on the xy-plane:

$$y = mx \implies z = \sqrt{x^2 + (mx)^2} = \sqrt{1 + m} \cdot |x|\,.$$

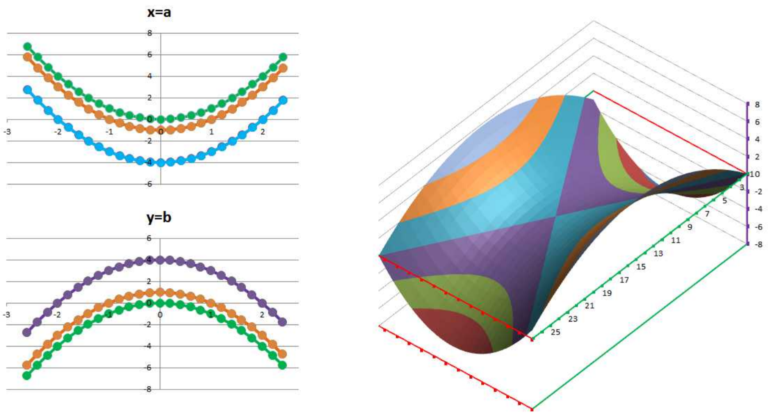
These are V-shaped curves.

Example 3.4.10: hyperbolas

Consider this function of two variables:

$$f(x, y) = y^2 - x^2\,.$$

Its level curves are hyperbolas again just as in the first example.



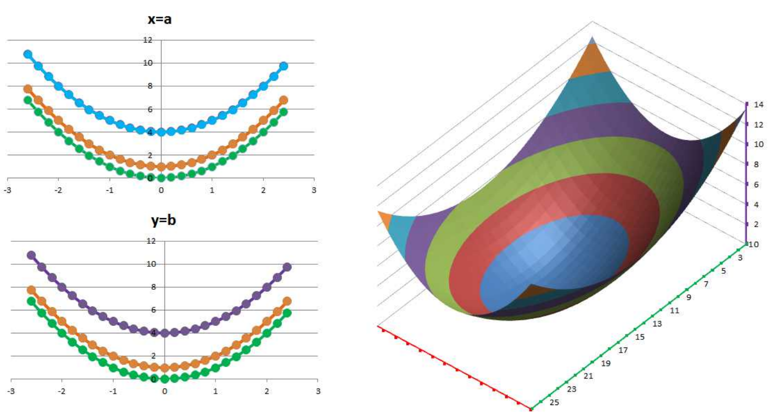
This is the familiar saddle point from the last section. Its variable curves are different simply because the angle of the graph relative to the axes is different.

Example 3.4.11: extreme point

If we replace $-$ with $+$, the function of two variables becomes

$$f(x,y) = y^2 + x^2$$

with a very different graph; it has an *extreme point*. The variable curves are still parabolas but both point upward this time:



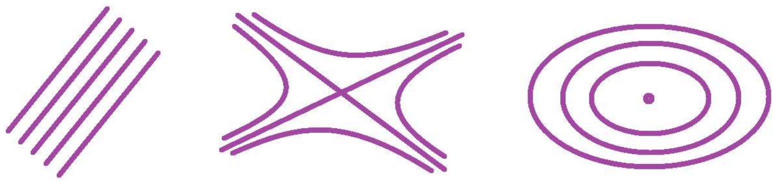
The level curves

$$y^2 + x^2 = c$$

are circles except when $c = 0$ (a point) or $c < 0$ (empty). They don't grow uniformly, as with the cone, with c however. The surface is called the *paraboloid of revolution*. One of the parabolas that passes through zero is rotated around the z-axis producing this surface.

The method of level curves has been used for centuries to create actual *maps*, i.e., 2-dimensional visualizations of 3-dimensional terrains:

The collection of all level curves is called the *contour map* of the function. We will see that zooming in on any point of the contour map is either a generic point with parallel level curves or a singular point exemplified by a saddle point and an extreme point.



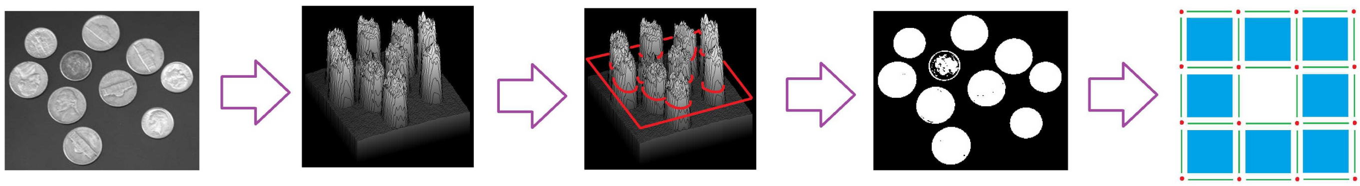
The singular points are little islands in the sea of generic points...

Example 3.4.12: sub-level set

Suppose $z = f(x, y)$ is a function of two variables, then for each value of $z = c$, the subset

$$\{(x, y) : \ f(x, y) \leq c\}$$

of the plane is called a *sub-level set* of f. These sets are used to convert gray-scale images to binary:



This operation is called "thresholding".

Thus, the variable curves are the result of *restricting the domain* of the function while the level curves are the result of *restricting the image*. Either method is applied in hope of simplifying the function to the degree that will make is subject to the tool we already have. However, the original graph is made of infinitely many of those...

Informally, the meaning of the domain is the same: the set of all possible inputs.

Definition 3.4.13: natural domain

The *natural domain* (or implied domain) of a function $z = f(X)$ is the set of all X in $\mathbf{R}^n$ for which $f(X)$ makes sense.

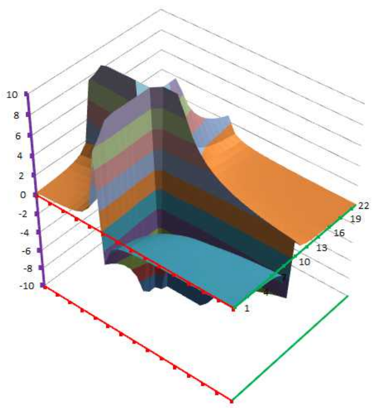
Just as before, the issue is that of division by zero, square roots, etc. The difference comes from the complexity of the space of inputs: the plane (and further $\mathbf{R}^n$) vs. the line.

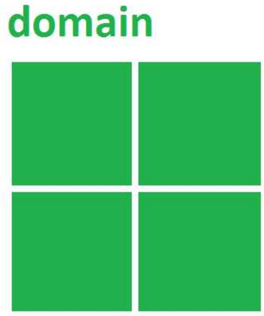
Example 3.4.14: reciprocal

The domain of the function

$$f(x, y) = \frac{1}{xy}$$

is the whole plane minus the axes:



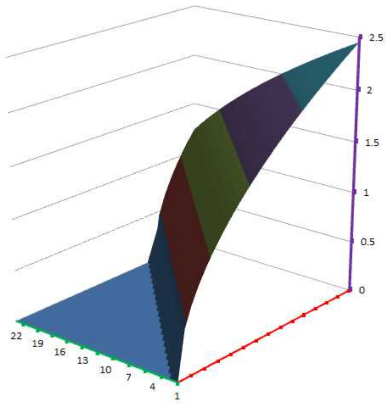


Example 3.4.15: root

The domain of the function

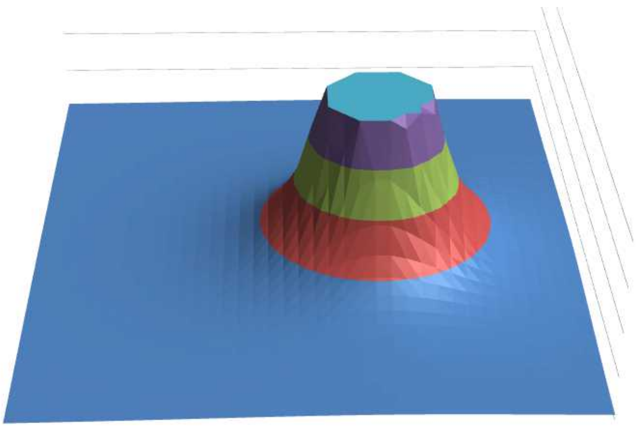
$$f(x,y)=\sqrt{x-y}$$

is the half of the plane given by the inequality $y \le x$:



Example 3.4.16: gravity

The domain of the function that represents the magnitude of the gravitational force is all points but 0:



It is a multiple of the function:

$$f(X)=\frac{1}{d(X,0)}\,.$$

Exercise 3.4.17

Provide the level and the variable curves for the function that represents the magnitude of the gravitational force.

Definition 3.4.18: graph of function

The *graph* of a function $z=f(X)$, where X belongs to $\mathbf{R}^n$, is the set of all points in $\mathbf{R}^{n+1}$ so that the first n coordinates are those of X and the last is $f(X)$.

We have used the restriction of the image and the level curves that come from it as a tool of *reduction*. We reduce the study of a complex object – the graph of $z=f(x,y)$ – to a collection of simpler ones – implicit curves $c=f(x,y)$.

The idea is even more useful in the higher dimensions. In fact, we can't simply plot the graph of a function of three variables anymore – it is located in $\mathbf{R}^4$ – as we did for functions of two variables. The level sets – *level surfaces* – is the best way to visualize it. We can also restrict the domain instead by fixing one variable at a time and plot the graphs of the *variable functions* of two variables.

The domains of functions of three variables are in our physical space. They may represent:

- the temperature or the humidity,
- the air or water pressure,
- the magnitude of a force (such as gravitation), etc.

Example 3.4.19: linear function

The level sets of a linear function

$$f(x, y, z) = A + mx + ny + kz\,.$$

are planes:

$$d = A + mx + ny + kz\,.$$

These planes located in xyz-space are all parallel to each other because they have the same normal vector $< m, n, k >$. The variable functions aren't that different; let's fix $z = c$:

$$d = f(x, y, c) = h(x, y) = A + mx + ny + kc\,.$$

For all values of c, these planes located in xyu-space are also parallel to each other because they have the same normal vector $< m, n, 1 >$.

Example 3.4.20: level sets

We start with a familiar function of *two* variables,

$$f(x, y) = \sin(xy)\,.$$

and just subtract z as the third producing a new function of *three* variables:

$$h(x, y, z) = \sin(xy) - z\,.$$

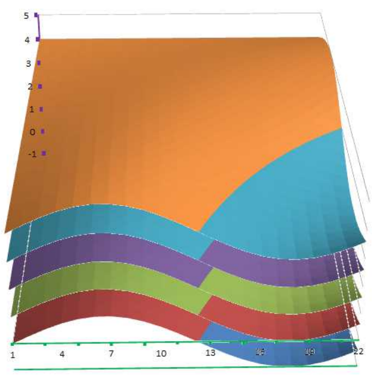
Then every of its level sets is given by:

$$d = \sin(xy) - z\,,$$

for some real d. What is this? We can make the relation explicit:

$$z = \sin(xy) - d\,,$$

Nothing but the graph of f shifted down (and up) by d:



In this sense, they are parallel to each other. The function is growing as we move in the direction of z. Now, if we fix an independent variable of h, say $z = c$, we have a function of two variables:

$$g(x, y) = \sin(xy) - c\,.$$

The graphs are the same.

Example 3.4.21: three variables

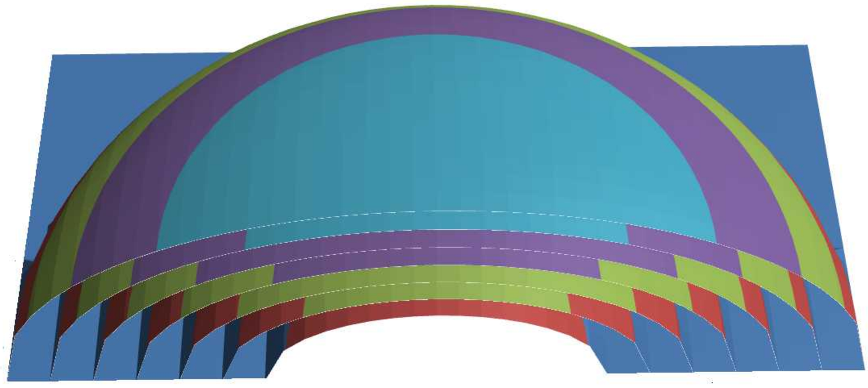
Let's consider this function of three variables:

$$f(x,y,z)=\sqrt{x^2+y^2+z^2}\,.$$

Then every of its level sets is given by this implicit equation:

$$d^2=x^2+y^2+z^2\,,$$

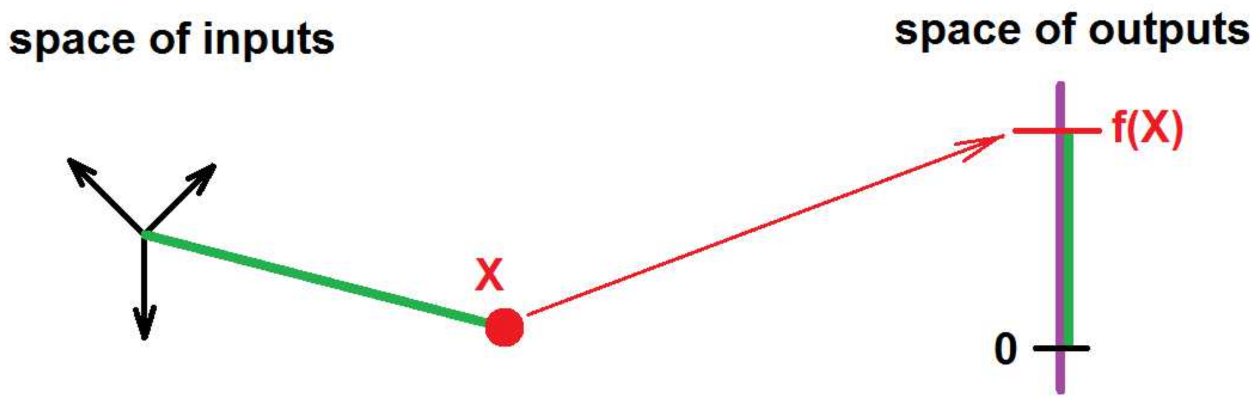
for some real d. Each is an equation of the sphere of radius $|d|$ centered at 0 (and the origin itself when $d=0$). They are concentric:



The radii also grow uniformly with d and, therefore, these surfaces are *not* are parallel to each other. So, the function is growing – and at a constant rate – as we move in any direction away from 0. What is this function? It's the *distance function* itself:

$$f(X)=f(x,y,z)=\sqrt{x^2+y^2+z^2}=||X||\,.$$

The graph of the magnitude of the gravitational force also has concentric spheres are level surfaces but the radii do not change uniformly with d. Both are visualized as follows:



The value of the function is decreasing as we move away from the center.

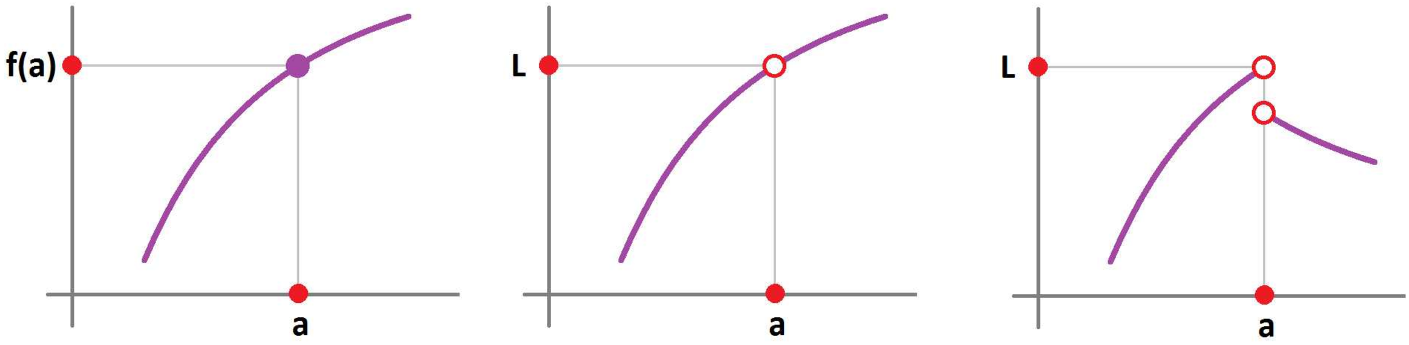
This is best we can do with functions of three variables. Beyond 3 variables, we are out of dimensions...

The approach via fixing the independent variables is considered later.

3.5. Limits

We now study small scale behavior of functions; we zoom in on a single point of the graph.

Just as in the lower dimensions, one of the most crucial properties of a function is the integrity of its graph: *is there a break or a cut or a hole?* For example, if we think of the graph as a terrain, is there a vertical drop?



We approach the issue via the *limits*.

In spite of all the differences between the functions we have seen – such as parametric curves vs. functions of several variables – the idea of limit is identical: as the input approaches a point, the output is also forced to approach a point. This is the context with the arrow "$\to$" to be read as "approaches":

numerical functions

$$y = f(x) \to l \quad \text{as } x \to a$$

parametric curves

$$Y = F(t) \to L \quad \text{as } t \to a$$

functions of several variables

$$z = f(X) \to l \quad \text{as } X \to A$$

We use lower case for scalars and upper case for anything multi-dimensional (point or vectors). We then see how the complexity shifts from the output to the input. And so does the challenge of multi-dimensionality.

We are ready for this challenge; this is the familiar meaning of convergence of a sequence in $\mathbf{R}^m$ to be used throughout:

$$X_n \to A \iff d(X_n, A) \to 0, \quad \text{or} \quad ||X_n - A|| \to 0\,.$$

The visualization is as expected:

The definition is almost an exact copy of what we used to have:

Definition 3.5.1: limit of a function

The *limit of a function* $z = f(A)$ at a point $X = A$ is defined to be the limit

$$\lim_{n\to\infty} f(X_n)$$

considered for all sequences $\{X_n\}$ within the domain of f excluding A that converge to A,

$$A \neq X_n \to A \quad \text{as } n \to \infty\,,$$

when all these limits exist and are equal to each other. In that case, we use the

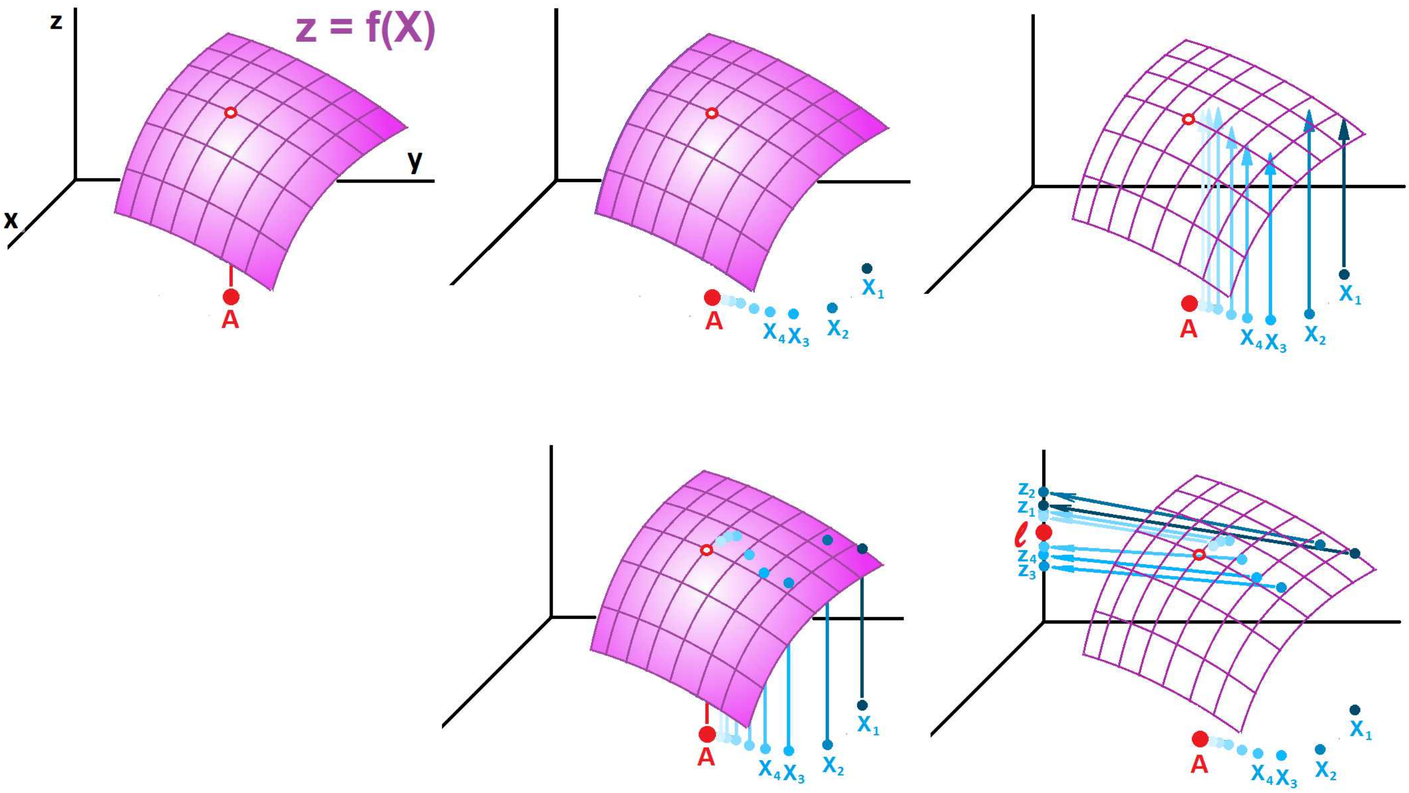
notation:

$$\lim_{X \to A} f(X)\,.$$

Otherwise, *the limit does not exist.*

We use this construction to study what is happening to $z = f(X)$ when a point X is in the vicinity of a chosen point $X = A$, where f might be undefined.

We start with an arbitrary sequence on the xy-plane that converges to this point, $X_n \to A$, then go vertically from each of these point to find the corresponding points on the graph of the function, $(X_n, f(X_n))$, and finally plot the output values (just numbers) on the z-axis, $z_n = f(X_n)$.

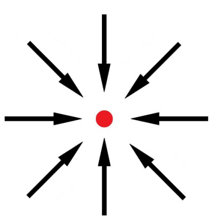


Is there a limit of this sequence? What about other sequences? Are all these limits the same?

The ability to approach the point from different directions had to be dealt with even for numerical functions:

$$\operatorname{sign}(x) \to -1 \ \text{ as } x \to 0^-, \text{ but } \operatorname{sign}(x) \to 1 \ \text{ as } x \to 0^+\,.$$

In the multi-dimensional case things are even more complicated as we can approach a point on the plane from infinitely many directions.



Example 3.5.2: no single plane

It is easy to come up with an example of a function that has different limits from different directions. We take one that will be important in the future:

$$\lim_{(x,y)\to(0,0)} \frac{|2x+y|}{\sqrt{x^2+y^2}}\,.$$

By the way, this is how one might try to calculate the "slope" (rise over the run) at 0 of the plane $z = 2x + y$. First, we approach along the x-axis, i.e., y is fixed at 0:

$$\lim_{x\to 0} \frac{|2x+0|}{\sqrt{x^2+0^2}} = \lim_{x\to 0} \frac{|2x|}{|x|} = 2\,.$$

Second, we approach along the y-axis, i.e., x is fixed at 0:

$$\lim_{y\to 0} \frac{|2\cdot 0+y|}{\sqrt{0^2+y^2}} = \lim_{y\to 0} \frac{|y|}{|y|} = 1\,.$$

There can't be just one slope for a plane...

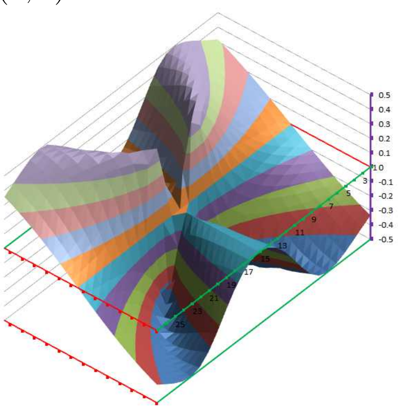
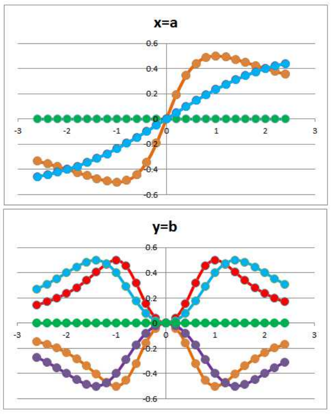
Things might be bad in an even more subtle way.

Example 3.5.3: smooth

Not all functions of two variables have graphs like the one above... Let's consider this function:

$$f(x,y) = \frac{x^2 y}{x^4+y^2}\,.$$

and its limit $(x,y)\to(0,0)$. Does it exist? It all depends on how X approaches $A = 0$:

The two green horizontal lines indicate that we get 0 if we approach 0 from either the x or the y direction. That's the horizontal cross we see on the surface (just like the one in the hyperbolic paraboloid). In fact, any (linear) direction is OK:

$$y = mx \implies f(x,mx) = \frac{x^2 mx}{x^4+(mx)^2} = \frac{mx^3}{x^4+m^2x^2} = \frac{m}{x+m^2/x} \to 0 \ \text{ as } x\to 0\,.$$

So, the limits from all directions are the same!

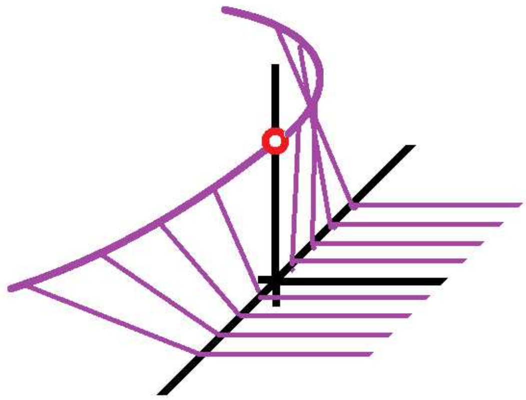
However, there seems to be two curved cliffs on left and right visible form above... What is we approach 0 along, instead of a straight line, a parabola $y = x^2$? The result is surprising:

$$y = x^2 \implies f(x,x^2) = \frac{x^2x^2}{x^4+(x^2)^2} = \frac{x^4}{2x^4} = \frac{1}{2}\,.$$

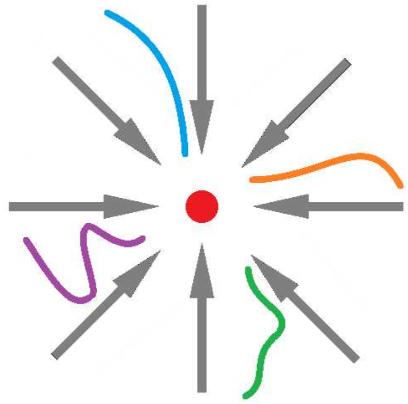
This is the height of the cliffs! So, this limit is different from the other and, according to our definition, the limit of the function does not exist:

$$\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4+y^2} \quad DNE\,.$$

In fact, the illustration created by the spreadsheet attempts to make an unbroken surface from these points while in reality there is no passage between the two cliffs as we just demonstrated.



So, not only we have to approach the point from all directions at once but also along any possible path.



A simpler but very important conclusion is that studying functions of several variables one variable at a time might be insufficient or even misleading.

A special note about the limit of a function at a point that lies on the *boundary* of the domain... It doesn't matter! The case of one-sided limits at a or b of the domain $[a,b]$ is now included in our new definition.

The algebraic theory of limits is almost identical to that for numerical functions. There are as many rules because whatever you can do with the outputs you can do with the functions.

We will use the algebraic properties of the limits of sequences – of numbers – to prove virtually identical facts about limits of functions.

Theorem 3.5.4: Algebra of Limits of Sequences

Suppose $a_n \to a$ and $b_n \to b$. Then:

SR:	$a_n + b_n \to a + b$	*CMR:*	$c \cdot a_n \to ca$	*real* c
PR:	$a_n \cdot b_n \to ab$	*QR:*	$a_n / b_n \to a/b$	$b \neq 0$

Each property is matched by its analog for functions.

Theorem 3.5.5: Algebra of Limits of Functions of Several Variables

Suppose $f(X) \to a$ and $g(X) \to b$ as $X \to A$. Then:

SR:	$f(X) + g(X) \to a + b$	*CMR:*	$c \cdot f(X) \to ca$	*real* c
PR:	$f(X) \cdot g(X) \to ab$	*QR:*	$f(X)/g(X) \to a/b$	$b \neq 0$

Note that there were no PR or QR for the parametric curves...

Let's consider them one by one.

Now, limits behave well with respect to the usual arithmetic operations.

Theorem 3.5.6: Sum Rule For Limits of Functions of Several Variables

If the limits at A of functions $f(X), g(X)$ exist then so does that of their sum, $f(P)+g(P)$, and the limit of the sum is equal to the sum of the limits:

$$\boxed{\lim_{X\to A}\big(f(X)+g(X)\big)=\lim_{X\to A}f(X)+\lim_{X\to A}g(X)}$$

Since the outputs and the limits are just numbers, in the case of infinite limits we follow the same rules of the algebra of infinities as in Volume 2 (Chapter 2DC-1):

$$\begin{array}{llll}\text{number} & + & (+\infty) & =+\infty\\ \text{number} & + & (-\infty) & =-\infty\\ +\infty & + & (+\infty) & =+\infty\\ -\infty & + & (-\infty) & =-\infty\end{array}$$

The proofs of the rest of the properties are identical.

Theorem 3.5.7: Constant Multiple Rule For Limits of Functions of Several Variables

If the limit at $X=A$ of function $f(X)$ exists then so does that of its multiple, $cf(X)$, and the limit of the multiple is equal to the multiple of the limit:

$$\boxed{\lim_{X\to A}cf(X)=c\cdot\lim_{X\to A}f(X)}$$

Theorem 3.5.8: Product Rule For Limits of Functions of Several Variables

If the limits at a of functions $f(X), g(X)$ exist then so does that of their product, $f(X)\cdot g(X)$, and the limit of the product is equal to the product of the limits:

$$\boxed{\lim_{X\to A}\big(f(X)\cdot g(X)\big)=\Big(\lim_{X\to A}f(X)\Big)\cdot\Big(\lim_{X\to A}g(X)\Big)}$$

Theorem 3.5.9: Quotient Rule For Limits of Functions of Several Variables

If the limits at $X=A$ of functions $f(X), g(X)$ exist then so does that of their ratio, $f(X)/g(X)$, provided $\lim\limits_{X\to A}g(X)\neq 0$, and the limit of the ratio is equal to the ratio of the limits:

$$\boxed{\lim_{X\to A}\left(\frac{f(X)}{g(X)}\right)=\frac{\lim\limits_{X\to A}f(X)}{\lim\limits_{X\to A}g(X)}}$$

Just as with sequences, we can represent these rules as commutative diagrams:

$$\begin{array}{ccc} f,g & \xrightarrow{\lim} & l,m \\ \downarrow + & SR & \downarrow + \\ f+g & \xrightarrow{\lim} & \lim(f+g)=l+m \end{array}$$

Exercise 3.5.10

Finish:

$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{x+y} = ?$$

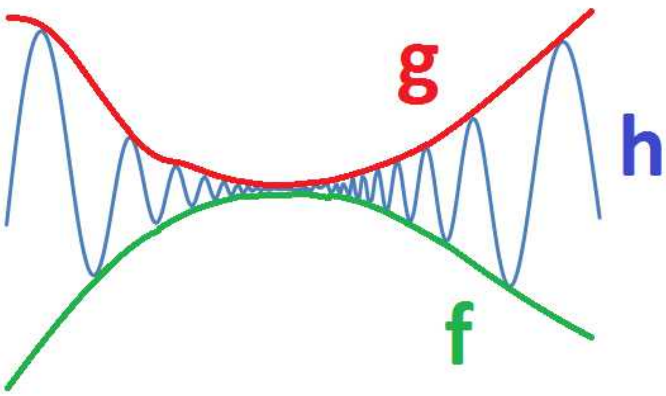
We can take advantage of the fact that the domain $\mathbf{R}^m$ of f can also be seen as made of *vectors* and vectors are subject to algebraic operations.

Theorem 3.5.11: Alternative Formula For Limits of Functions of Several Variables

The limit of a function $z=f(X)$ at $X=A$ is equal to l if and only if

$$\lim_{||H||\to 0} f(A+H) = l\,.$$

The next result stands virtually unchanged from Volume 2 (Chapter 2DC-2).



Theorem 3.5.12: Squeeze Theorem For Functions of Several Variables

If a function is squeezed between two functions with the same limit at a point, its limit also exists and is equal to the that number; i.e., if

$$f(X) \le h(X) \le g(X)\,,$$

for all X within some distance from $X=A$, and

$$\lim_{X\to A} f(X) = \lim_{X\to A} g(X) = l\,,$$

then the following limit exists and equal to that number:

$$\lim_{X\to A} h(X) = l\,.$$

The easiest way to handle limits is *coordinatewise*, when possible.

Theorem 3.5.13: Limit of Function of Several Variables

If the limit of a function of several variables exists then it exists with respect to each of the variables; i.e., for any function $z = f(X) = f(x_1, ..., x_m)$ *and any* $A = (a_1, ..., a_m)$ *in* $\mathbf{R}^m$, *we have:*

$$f(X) \to l \ \text{ as } X \to A \\ \Longrightarrow \ f(a_1, ..., a_{k-1}, x, a_{k+1}, ..., a_m) \to l \ \text{ as } x \to a_k \ ,$$

for each $k = 1, 2, ..., m$.

The converse isn't true!

Example 3.5.14: 2d limit

Recall that this function has limits at $(0, 0)$ with respect to either of the variables:

$$f(x, y) = \frac{x^2 y}{x^4 + y^2} ,$$

but the limit as $(x, y) \to (0, 0)$ does not exist.

So, we have to establish that the limit exists – in the "omni-directional" sense – first and only then we can use limits with respect to every variable – in the "uni-directional" sense – to find this limit.

Now, the asymptotic behavior...

Definition 3.5.15: function approaches infinity

Given a function $z = f(X)$ and a point A in $\mathbf{R}^n$, we say that f *approaches infinity at* A if

$$\lim_{n\to\infty} f(x) = \pm\infty ,$$

for any sequence $X_n \to A$ as $n \to \infty$. Then we use the notation:

$$\lim_{X\to A} f(x) = \pm\infty .$$

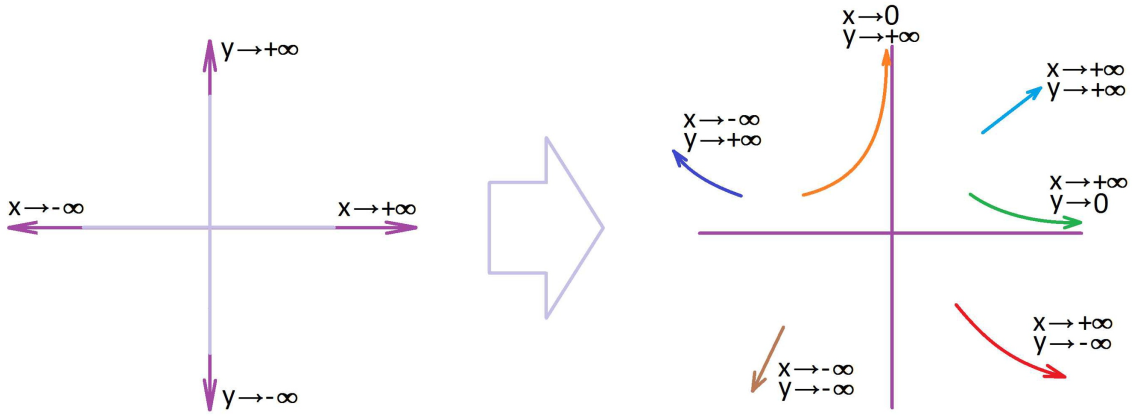
The line $X = A$ located in $\mathbf{R}^{n+1}$ is then called a *vertical asymptote* of f.

Example 3.5.16: Newton's Law of Gravity

If $z = f(x, y, z)$ represents the magnitude of the force of gravity of an object located at the point $X = (x, y, z)$ relative to another object located at the origin, then we have:

$$\lim_{X\to 0} f(X) = \infty .$$

Next, we can approach infinity in a number of ways too:



That's why we have to look at the distance to the origin.

Definition 3.5.17: function of several variables goes to infinity

For a function of several variables $z = f(X)$, we say that f *goes to infinity* if

$$f(X_n) \to \pm\infty\,,$$

for any sequence X_n with $||X_n|| \to \infty$ as $n \to \infty$. Then we use the notation:

$$f(X) \to \pm\infty \quad \text{as } X \to \infty\,,$$

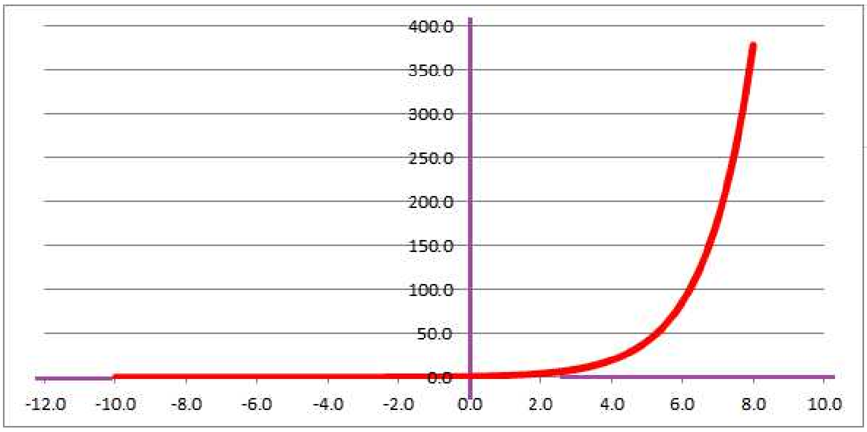
or

$$\lim_{x\to\infty} f(x) = \pm\infty\,.$$

Example 3.5.18: horizontal asymptotes

We previously demonstrated the following:

$$\lim_{x\to-\infty} e^x = 0, \qquad \lim_{x\to+\infty} e^x = +\infty\,.$$



We won't speak of horizontal asymptotes but the next idea is related.

Definition 3.5.19: function approaches value at infinity

Given a function $z = f(X)$, we say that f *approaches* $z = d$ *at infinity* if

$$\lim_{n\to\infty} f(X_n) = d\,,$$

for any sequence X_n with $||X_n|| \to \infty$ as $n \to \infty$. Then we use the notation:

$$\lim_{X\to\infty} f(X) = d\,.$$

Example 3.5.20: Newton's Law of Gravity

If $z = f(x, y, z)$ represents the magnitude of the force of gravity of an object located at the point $X = (x, y, z)$ relative to another object located at the origin, then we have:

$$\lim_{X\to\infty} f(X) = 0\,.$$

3.6. Continuity

The idea of continuity is identical to that for numerical functions or parametric cures: as the input approaches a point, the output is forced to approach the values of the function at that point. The concept flows from the idea of limit just as before.

Definition 3.6.1: function continuous at point

A function $z = f(X)$ is called *continuous at point* $X = A$ when:

- $f(X)$ is defined at $X = A$.
- The limit of f exists at A.
- The two are equal to each other:

$$\lim_{X\to A} f(X) = f(A)\,.$$

Furthermore, a function is *continuous* if it is continuous at every point of its domain.

Thus, the limits of continuous functions can be found by *substitution*:

$$\boxed{\lim_{X\to A} f(X) = f(A)}$$

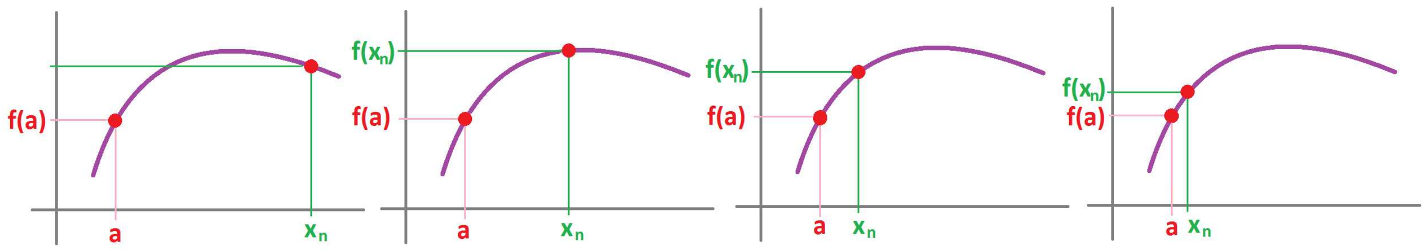
or

$$\boxed{f(X) \to f(A) \quad \text{as } X \to A}$$

Equivalently, a function f is continuous at a if we have:

$$\lim_{n\to\infty} f(A_n) = f(A)\,,$$

for any sequence $X_n \to A$.



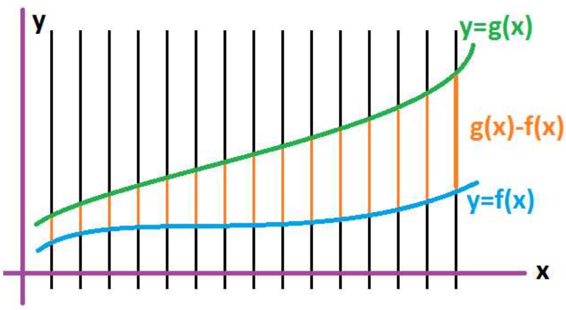
A typical function we encounter is continuous at every point of its domain.

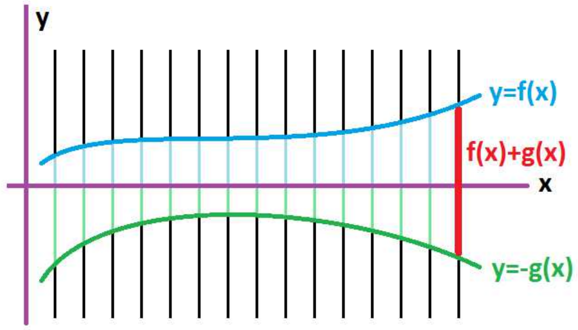
Theorem 3.6.2: Algebra of Continuity

Suppose f and g are continuous at $X = A$. Then so are the following functions:

1. *(SR) $f \pm g$,*
2. *(CMR) $c \cdot f$ for any real c,*
3. *(PR) $f \cdot g$, and*
4. *(QR) f/g provided $g(A) \neq 0$.*

As an illustration, we can say that if the floor and the ceiling represented by f and g respectively of a cave are changing continuously then so is its height, which is $g - f$:





Or, if the floor and the ceiling (f and $-g$) are changing continuously then so is its height ($g + f$). And so on.

There are two ways to compose a function of several variables with another function...

Theorem 3.6.3: Composition Rule I

If the limit at $X = A$ of function $z = f(X)$ exists and is equal to l then so does that of its composition with any numerical function $u = g(z)$ continuous at $z = l$ and

$$\boxed{\lim_{X \to A} (g \circ f)(X) = g(l)}$$

Proof.

Suppose we have a sequence,

$$X_n \to A\,.$$

Then, we also have another sequence,

$$b_n = f(X_n)\,.$$

The condition $f(X) \to l$ as $X \to A$ is restated as follows:

$$b_n \to l \quad \text{as } n \to \infty\,.$$

Therefore, continuity of g implies,

$$g(b_n) \to g(l) \quad \text{as } n \to \infty\,.$$

In other words,

$$(g \circ f)(X_n) = g(f(X_n)) \to g(l) \quad \text{as } n \to \infty\,.$$

Since sequence $X_n \to A$ was chosen arbitrarily, this condition is restated as,

$$(g \circ f)(X) \to g(l) \quad \text{as } X \to A\,.$$

We can rewrite the result as follows:

$$\boxed{\lim_{X\to A}(g\circ f)(X) = g(l)\Big|_{l=\lim_{X\to A} f(X)}}$$

Furthermore, we have:

$$\boxed{\lim_{X\to A} g\big(f(X)\big) = g\left(\lim_{X\to A} f(X)\right)}$$

Theorem 3.6.4: Composition Rule II

If the limit at $t=a$ of a parametric curve $X=F(t)$ exists and is equal to L then so does that of its composition with any function $z=g(X)$ of several variables continuous at $X=L$ and

$$\boxed{\lim_{t\to a}(g\circ F)(t) = g(L)}$$

Proof.

Suppose we have a sequence,

$$t_n \to a\,.$$

Then, we also have another sequence,

$$B_n = F(t_n)\,.$$

The condition $F(t)\to L$ as $t\to a$ is restated as follows:

$$B_n \to L \quad \text{as } n\to\infty\,.$$

Therefore, continuity of F implies,

$$g(B_n)\to g(L) \quad \text{as } n\to\infty\,.$$

In other words,

$$(g\circ F)(t_n) = g(F(t_n)) \to g(L) \quad \text{as } n\to\infty\,.$$

Since sequence $X_n\to A$ was chosen arbitrarily, this condition is restated as,

$$(g\circ F)(t)\to g(L) \quad \text{as } t\to a\,.$$

We can rewrite the result as follows:

$$\boxed{\lim_{t\to a}(g\circ F)(t) = g(L)\Big|_{L=\lim_{t\to a} F(t)}}$$

Furthermore, we have:

$$\boxed{\lim_{t\to a} g\big(F(t)\big) = g\left(\lim_{t\to a} F(t)\right)}$$

Corollary 3.6.5: Composition of Continuous Functions

The composition $g \circ f$ of a function f continuous at $x = a$ and a function g continuous at $y = f(a)$ is continuous at $x = a$.

The easiest way to handle continuity is *coordinatewise*, when possible.

Theorem 3.6.6: Continuity of Function of Several Variables

If a function of several variables is continuous then it is also continuous with respect to each of the variables.

The converse isn't true!

Example 3.6.7: 2d limit

Recall that this function is continuous at $(0,0)$ with respect to either of the variables:

$$f(x,y) = \frac{x^2 y}{x^4 + y^2}\,,$$

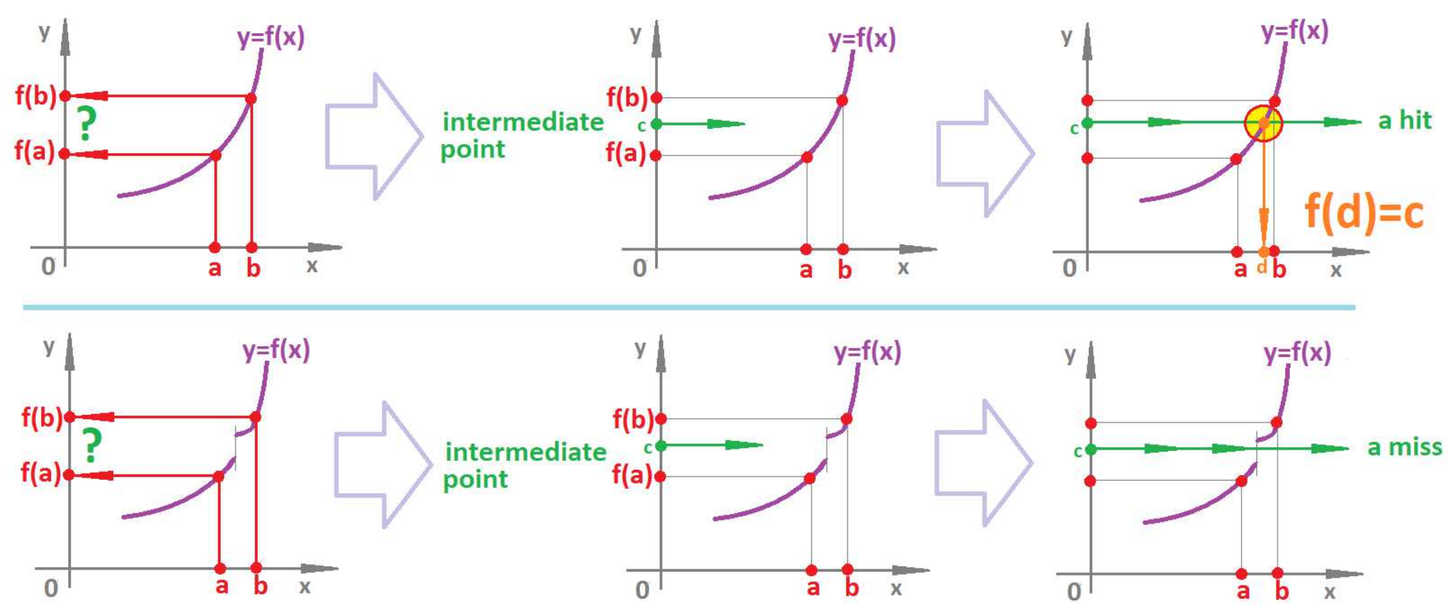
but the limit as $(x,y) \to (0,0)$ simply does not exist.

So, we have to establish continuity – in the sense of the "omni-directional" limit – first and only then we can use this fact to find facts about the continuity with respect to every variable – in the "uni-directional" sense.

A special note about the continuity of a function at a point that lies on the *boundary* of the domain... Once again, it doesn't matter! The case of one-sided continuity at a or b of the domain $[a,b]$ is now included in our new definition.

The definition of continuity is purely *local*: only the behavior of the function in the, no matter how small, vicinity of the point matters. If the function is continuous on a whole set, what can we say about its *global* behavior?

Our understanding of continuity of numerical functions has been as the property of having *no gaps in their graphs*.



This idea is more precisely expressed by the following by the *Intermediate Value Theorem*:

▶ If a function f is defined and is continuous on an interval $[a,b]$, then for any c between $f(a)$ and $f(b)$, there is d in $[a,b]$ such that $f(d) = c$.

An often more convenient way to state is:

- ▶ If the domain of a continuous function is an interval then so is its image.

Now the plane is more complex than a line and we can't limit ourselves to intervals. But what is the analog of an interval in a multidimensional space?

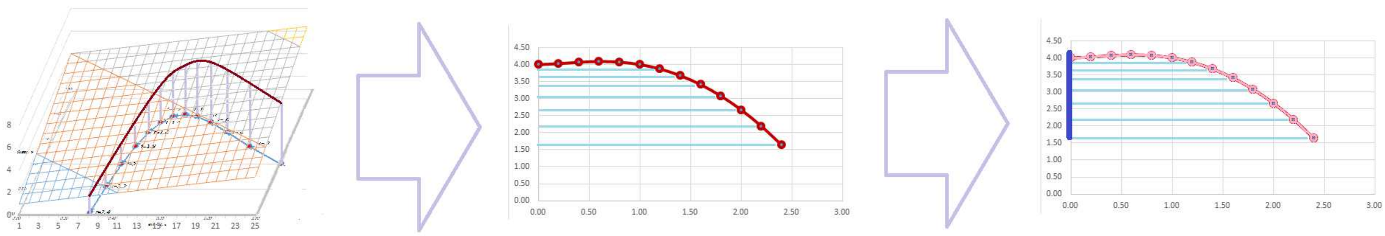
Theorem 3.6.8: Intermediate Value Theorem for Functions of Several Variables

Suppose a function f is defined and is continuous on a set that contains the path C of a continuous parametric curve. Then the image of this path is an interval.

Proof.

It follows from the *Composition Rule II* and the *Intermediate Value Theorem* for numerical functions.

The theorem says that there are no missing values in the image of such a set.



Exercise 3.6.9

Show that that the converse of the theorem isn't true.

A convenient re-statement of the theorem is below.

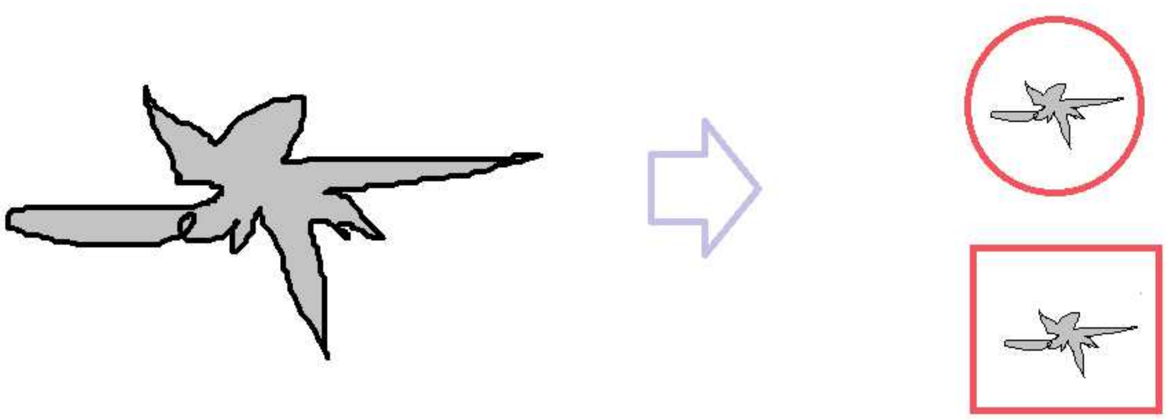
Corollary 3.6.10: Image of Continuous Function

The image of a path-connected set under a continuous function is an interval.

Recall that a function f is called *bounded* on a set S in $\mathbf{R}^n$ if its image is bounded, i.e., there is such a real number m that

$$|f(X)| \leq m$$

for all X in S.



Theorem 3.6.11: Cont. => Bounded

If the limit at $X = A$ of function $z = f(X)$ exists then f is bounded on some

open disk that contains A:

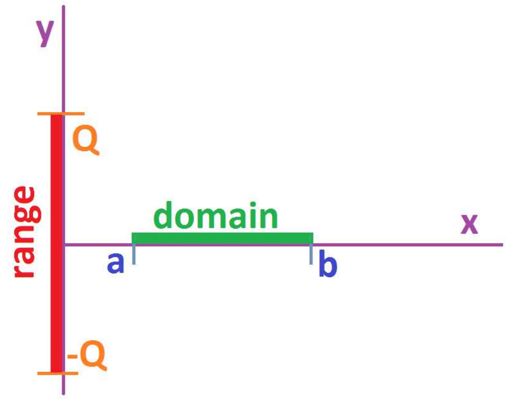
$$\lim_{X\to A} f(X) \text{ exists} \implies |f(X)| \le m$$

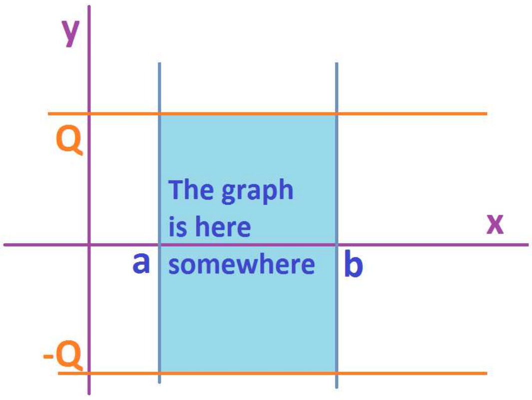
for all X with $d(X,A) < \delta$ for some $\delta > 0$ and some m.

Exercise 3.6.12

Prove the theorem.

The *global* version of the above theorem guarantees that the function is bounded under certain circumstances. The version of the theorem for numerical functions is simply: a continuous on $[a,b]$ function is bounded.





But what is the multi-dimensional analog of a closed bounded interval?

We already know that a set S in $\mathbf{R}^n$ is *bounded* if it fits in a sphere (or a box) of a large enough size:

$$||x|| < Q \text{ for all } x \text{ in } S\,;$$

and a set in $\mathbf{R}^n$ is called *closed* if it contains the limits of all of its convergent sequences.

Theorem 3.6.13: Boundedness

A continuous on a closed bounded set function is bounded.

Proof.

Suppose, to the contrary, that $x = f(X)$ is unbounded on set S. Then there is a sequence $\{X_n\}$ in S such that $f(X_n) \to \infty$. Then, by the *Bolzano-Weierstrass Theorem*, sequence $\{X_n\}$ has a convergent subsequence $\{Y_k\}$:

$$Y_k \to Y\,.$$

This point belong to S! From the continuity, it follows that

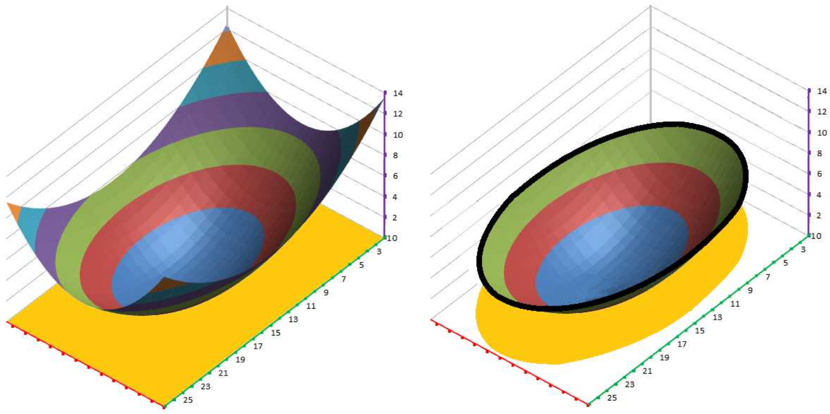
$$f(Y_k) \to f(Y)\,.$$

This contradicts the fact that $\{Y_k\}$ is a subsequence of a sequence that diverges to ∞.

Exercise 3.6.14

Why are we justified to conclude in the proof that the limit Y of $\{Y_k\}$ is in S?

Is the image of a *closed* set closed? If it is, the function *reaches its extreme values*, i.e., the least upper bound sup and the greatest lower bound inf.



Definition 3.6.15: global maximum and minimum points

Given a function $z = f(X)$. Then $X = D$ is called a *global maximum point* of f on set S if

$$f(D) \ge f(X) \ \text{ for all } X \text{ in } S\,.$$

And $X = C$ is called a *global minimum point* of f on set S if

$$f(C) \le f(X) \ \text{ for all } X \text{ in } S\,.$$

(They are also called *absolute maximum and minimum points.*) Collectively they are all called *global extreme points.*

Just because something is described doesn't mean that it can be found.

Theorem 3.6.16: Extreme Value Theorem

A continuous function on a closed bounded set in $\mathbf{R}^n$ attains its global maximum and global minimum values; i.e., if $z = f(X)$ is continuous on a bounded closed set S, then there are C, D in S such that

$$f(C) \le f(X) \le f(D)\,,$$

for all X in S.

Proof.

It follows from the *Bolzano-Weierstrass Theorem.*

Definition 3.6.17: global maximum and minimum value

Given a function $z = f(X)$. Then $X = M$ is called the *global maximum value* of f on set S if

$$M \ge f(X) \ \text{ for all } X \text{ in } S\,.$$

And $y = m$ is called the *global minimum value* of f on set S if

$$m \le f(X) \ \text{ for all } X \text{ in } S\,.$$

(They are also called *absolute maximum and minimum values.*) Collectively they are all called *global extreme values.*

Then *the* global max (or min) value is reached by the function at *any* of its global max (or min) points.

Note that the reason we need the *Extreme Value Theorem* is to ensure that the optimization problem we are facing has a solution.

We can define limits and continuity of functions of several variables without invoking limits of sequences. Let's rewrite what we want to say about the meaning of the limits in progressively more and more precise terms.

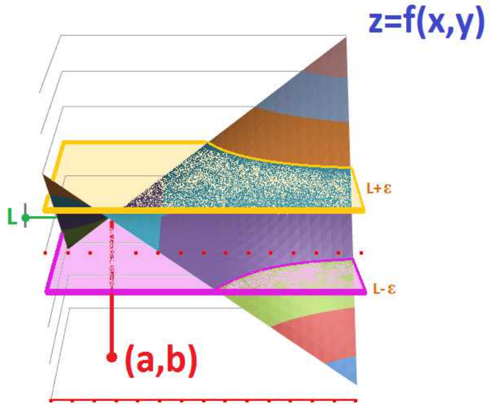
X	$z = f(X)$
As $X \to A$,	we have $y \to l$.
As X approaches A,	y approaches l.
As the distance from X to A approaches 0,	the distance from y to l approaches 0.
As $d(X, A) \to 0$,	we have $\lvert y - l\rvert \to 0$.
By making $d(X, A)$ as smaller and smaller,	we make $\lvert y - l\rvert$ as small as needed.
By making $d(X, A)$ less than some $\delta > 0$,	we make $\lvert y - l\rvert$ smaller than any given $\varepsilon > 0$.

Definition 3.6.18: limit of function

The *limit of function* $z = f(X)$ at $X = A$ is a number l, if exists, such that for any $\varepsilon > 0$ there is such a $\delta > 0$ that

$$0 < d(X, A) < \delta \implies |f(X) - l| < \varepsilon\,.$$

This is the geometric meaning of the definition: if X is within δ from A, then $f(X)$ is supposed to be within ε from l. In other words, this part of the graph fits between the two planes ε away from the *plane* $z = l$.



3.7. The difference and the partial difference quotients

We start with *numerical* differentiation of functions of several variables.

Example 3.7.1: hyperbolic paraboloid

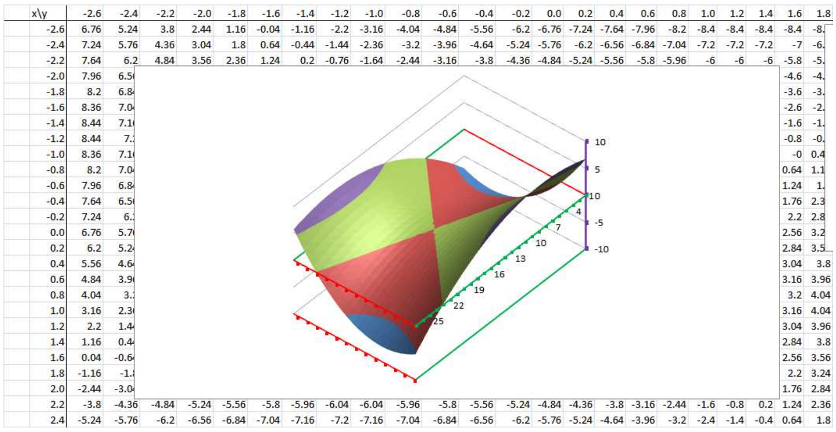
We consider the function:

$$f(x, y) = -x^2 + y^2 + xy\,.$$

For each x in the left-most column and each y in the top row, the corresponding value of the function is computed and placed in this table, just as before:

```
=R2C3*(R4C^2-RC2^2+R4C*RC2)
```

When plotted is recognized as a familiar hyperbolic paraboloid:



Below we outline the process of *partial differentiation*. The variable functions – with respect to x and y – are shown. These are the functions we will differentiate.

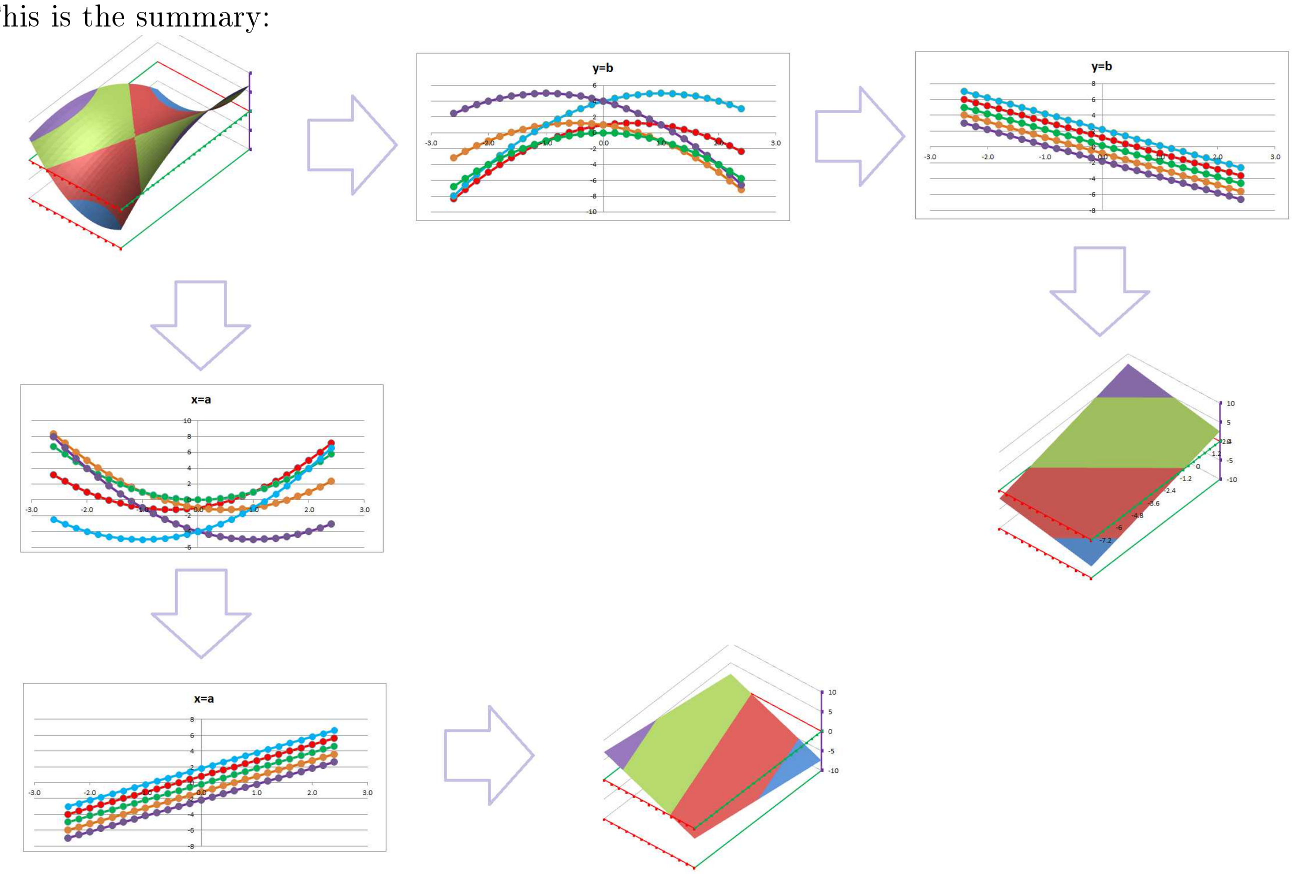
First, for each value of y given in the top row, we compute the difference quotient function with respect to x by going down the corresponding column and then placing these values on right in a new table (it is one row short in comparison to the original):

```
=(RC[-29]-R[-1]C[-29])/R2C1
```

Second, for each value of x given in the left-most column, we compute the difference quotient function with respect to y by going right the corresponding row and then placing these values below in a new table (it is one column short in comparison to the original).

```
=(R[-29]C-R[-29]C[-1])/R2C1
```

This is the summary:



The results are all straight lines equally spaced and in both cases they form planes. These planes are the graphs of the two new functions of two variables, the *partial derivatives*, the tables of which have been constructed. Some things don't change: just as in the one-dimensional case, *the derivatives of quadratic functions are linear.*

We deal with the change of the values of a function, Δf, relative to the change of its input variable. This time, there are two: Δx and Δy.

If we know only four values of a function of two variables (left), we can compute the *differences* $\Delta_x f$ and $\Delta_y f$ of f along both horizontal and vertical edges (right):

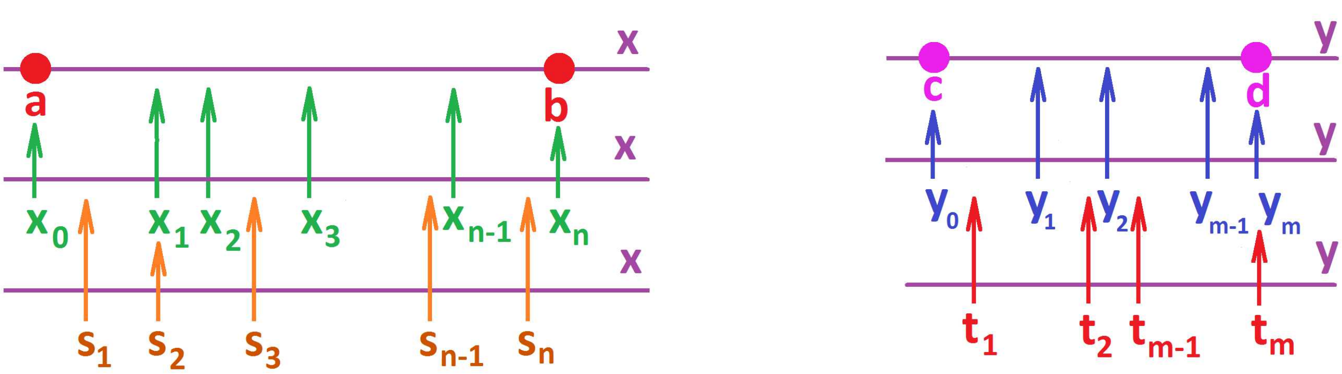
$$\begin{array}{rccccccc}
y+\Delta y: & f(x,y+\Delta y) & --- & f(x+\Delta x,y+\Delta y) & & -\bullet- & \Delta_x f(s,y+\Delta y) & -\bullet- \\
 & | & & | & & | & & | \\
t: & | & & | & \to & \Delta_y f(x,t) & & \Delta_y f(x+\Delta x,t) \\
 & | & & | & & | & & | \\
y: & f(x,y) & --- & f(x+\Delta x,y) & & -\bullet- & \Delta_x f(s,y) & -\bullet- \\
 & x & s & x+\Delta x & & x & s & x+\Delta x
\end{array}$$

As you can see, we subtract the values at the corners – vertically and horizontally – and place them at the corresponding edges.

We then acquire the *difference quotients* by dividing by the increment of the corresponding variable:

$$\begin{array}{cccc}
 & -\bullet- & \dfrac{\Delta f}{\Delta x}(s,y+\Delta y) & -\bullet- \\
 & | & & | \\
\to & \dfrac{\Delta f}{\Delta y}(x,t) & & \dfrac{\Delta f}{\Delta y}(x+\Delta x,t) \\
 & | & & | \\
 & -\bullet- & \dfrac{\Delta f}{\Delta x}(s,y) & -\bullet- \\
 & x & s & x+\Delta x
\end{array}$$

More generally, we build a *partition P of a rectangle* $R=[a,b]\times[c,d]$ in the xy-plane. It is a combination of partitions of the intervals $[a,b]$ and $[c,d]$:



We start with a partition of an interval $[a,b]$ in the x-axis into n intervals:

$$[x_0,x_1],\ [x_1,x_2],\ \dots,\ [x_{n-1},x_n]\,,$$

with $x_0=a,\ x_n=b$. The increments of x are:

$$\Delta x_i = x_i - x_{i-1},\ i=1,2,\dots,n\,.$$

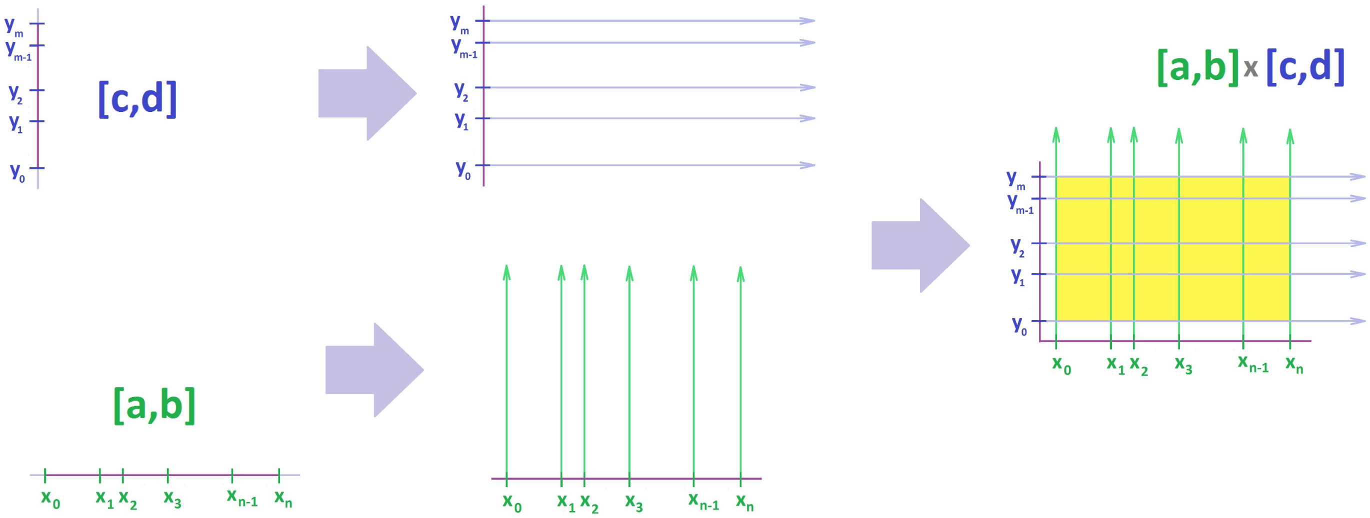
Then we do the same for y. We partition an interval $[c,d]$ in the y-axis into m intervals:

$$[y_0,y_1],\ [y_1,y_2],\ \dots,\ [y_{m-1},y_m]\,,$$

with $y_0=c,\ y_n=d$. The increments of y are:

$$\Delta y_i = y_i - y_{i-1},\ i=1,2,\dots,m\,.$$

The lines $y = y_j$ and $x = x_i$ create a partition P of the rectangle $[a, b] \times [c, d]$ into smaller rectangles $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$:



The points of intersection of these lines,

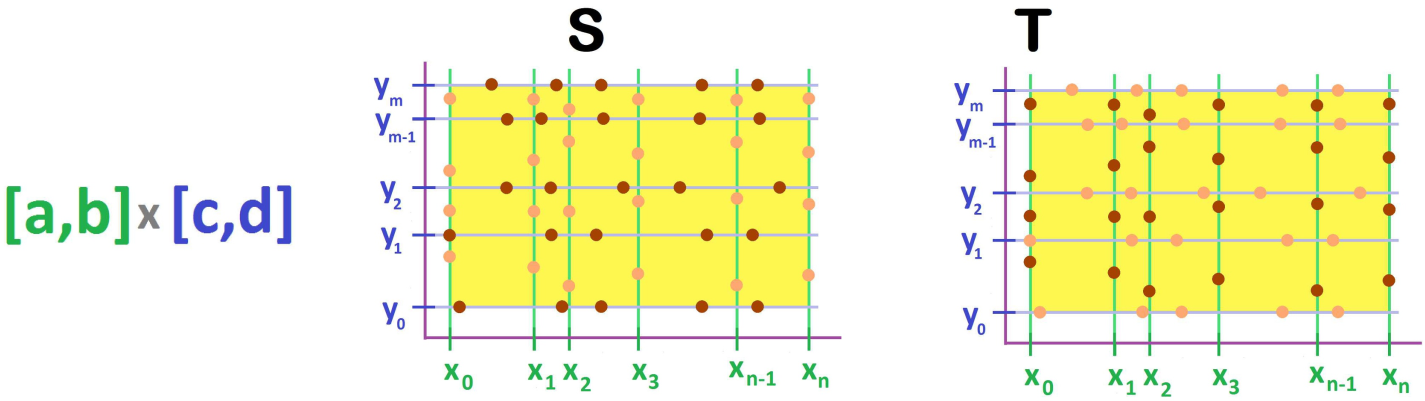
$$X_{ij} = (x_i, y_j),\ i = 1, 2, ..., n,\ j = 1, 2, ..., m\,,$$

will be called the (primary) *nodes* of the partition.

The *secondary nodes* of P appear on each of the horizontal and vertical edges of the partition; for each pair $i = 0, 1, 2, ..., n-1$ and $i = 0, 1, 2, ..., m-1$, we have:

- a point S_{ij} in the segment $[x_i, x_{i+1}] \times \{y_j\}$, and
- a point T_{ij} in the segment $\{x_i\} \times [y_j, y_{j+1}]$.

The locations within the segments are arbitrary:

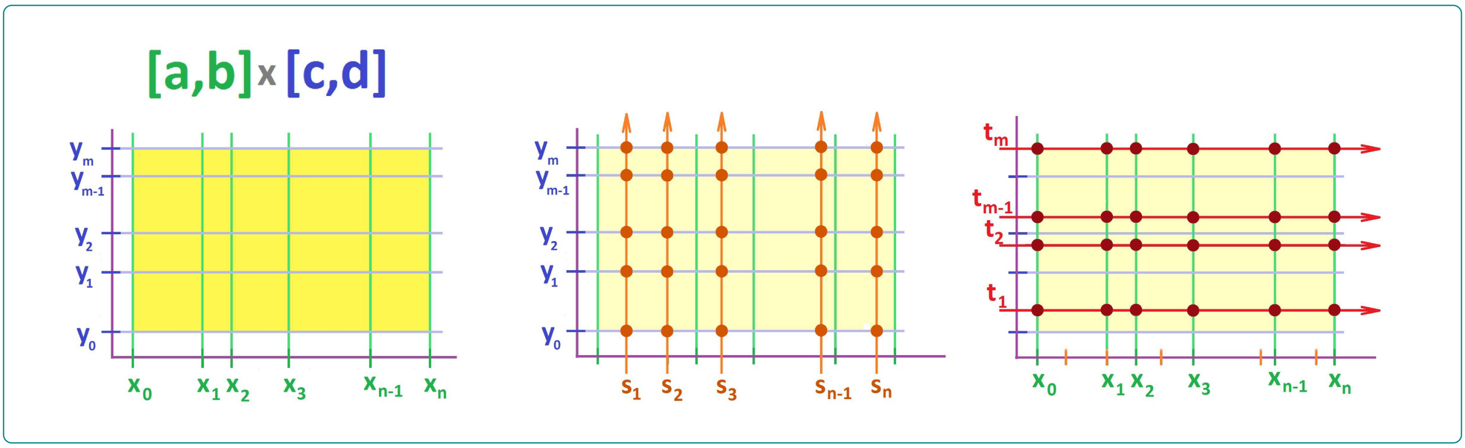


We can have left- and right-end augmented partitions as well as mid-point ones.

Example 3.7.2: augmented partition

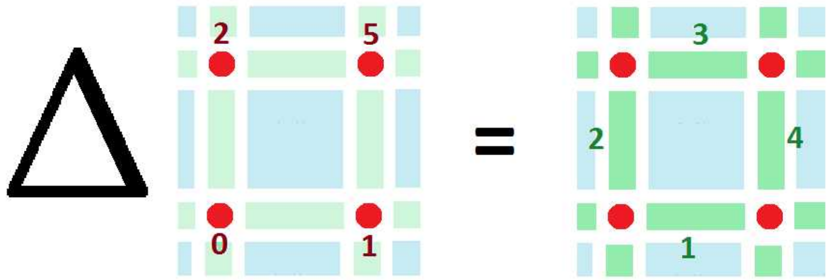
As a simple choice, we can use the secondary nodes of the augmented partitions of $[a, b]$, say $\{s_i\}$, and $[c, d]$, say $\{t_j\}$:

- a point $S_{ij} = (s_i, y_j)$ in the segment $[x_{i-1}, x_i] \times \{y_j\}$, and
- a point $T_{ij} = (x_i, t_j)$ in the segment $\{x_i\} \times [y_{j-1}, y_j]$.



Now calculus.

Suppose we have a function known only at the nodes (left):



When $y = y_j$ is fixed, its difference Δf is computed over each interval of the partition $[x_i, x_{i+1}]$, $i = 0, 1, 2..., n-1$ of the segment (right). This defines a new function on the secondary nodes. Similar for every fixed $x = x_i$.

It is a 1-form. It includes both horizontal and vertical differences.

The definition of the difference is identical to the ones we have seen. Suppose a function of two variables $z = f(X)$ is defined at the nodes of a partition of a region.

Definition 3.7.3: difference

The *difference* f is defined at the secondary node of the edge $N = [X, X + \Delta X]$ of the partition by:

$$\boxed{\Delta f\,(N) = f(X + \Delta X) - f(X)}$$

The definition work for all dimensions.

Often, we prefer to be more specific about the increments of X. Suppose f is defined at the nodes X_{ij}, $i, j = 0, 1, 2, ..., n$, of the partition. Then we follow the two axes separately:

Definition 3.7.4: partial difference

The *partial difference of f with respect to x* is defined at the secondary nodes of the partition by:

$$\boxed{\Delta_x f\,(S_{ij}) = f(X_{ij}) - f(X_{i-1,j})}$$

and *partial difference of f with respect to y* is defined at the secondary nodes of the partition by:

$$\boxed{\Delta_y f\,(T_{ij}) = f(X_{ij}) - f(X_{i,j-1})}$$

The connection is obvious:

Theorem 3.7.5: Difference And Partial Differences

The difference of $z = f(x, y)$ is the function defined at each secondary node of the partition and is equal to the corresponding partial difference of f with respect to x or y, denoted as follows:

$$\boxed{\Delta f\,(N) = \begin{cases} \Delta_x f\,(S_{ij}) & \text{if } N = S_{ij}\,, \\ \Delta_y f\,(T_{ij}) & \text{if } N = T_{ij}\,. \end{cases}}$$

This is our true interest as explained above:

Definition 3.7.6: partial difference quotient

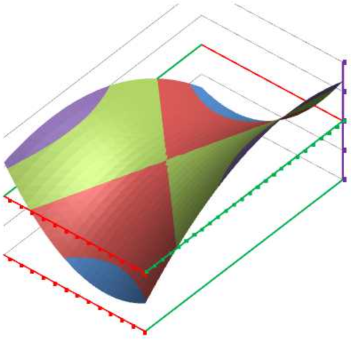
The *partial difference quotient of f with respect to x* is defined at the secondary nodes of the partition by:

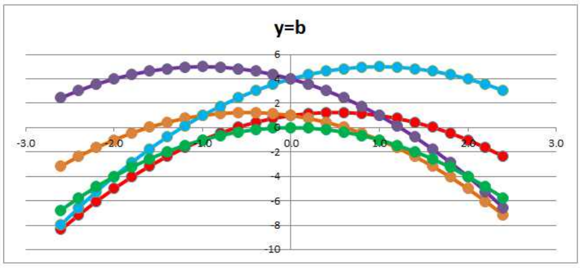
$$\boxed{\frac{\Delta f}{\Delta x}(S_{ij}) = \frac{\Delta_x f\,(S_{ij})}{\Delta x_i} = \frac{f(X_{ij}) - f(X_{i-1,j})}{x_i - x_{i-1}}}$$

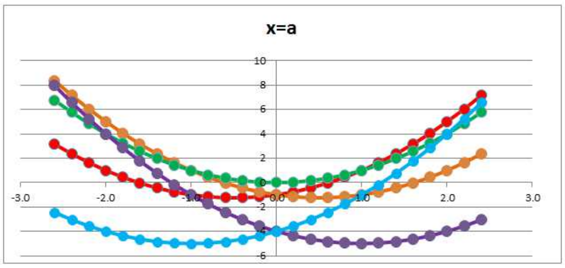
And *partial difference quotient of f with respect to y* is defined at the secondary nodes of the partition by:

$$\boxed{\frac{\Delta f}{\Delta y}(T_{ij}) = \frac{\Delta_y f\,(T_{ij})}{\Delta y_j} = \frac{f(X_{ij}) - f(X_{i,j-1})}{y_j - y_{j-1}}}$$

These two numbers represent the slopes of the secant lines along the x-axis and the y-axis over the nodes respectively:







Note that both $\frac{\Delta f}{\Delta x}$ and $\frac{\Delta f}{\Delta y}$ are literally fractions.

For each pair (i, j), the two difference quotients appear as the coefficients in the equation of the plane

through the three points on the graph above the three adjacent nodes in the xy-plane:

$$\begin{array}{llllll} (& x_i, & y_j, & f(x_i, & y_j) &) \\ (& x_{i+1}, & y_j, & f(x_{i+1}, & y_j) &) \\ (& x_i, & y_{j+1}, & f(x_i, & y_{j+1}) &) \end{array}$$

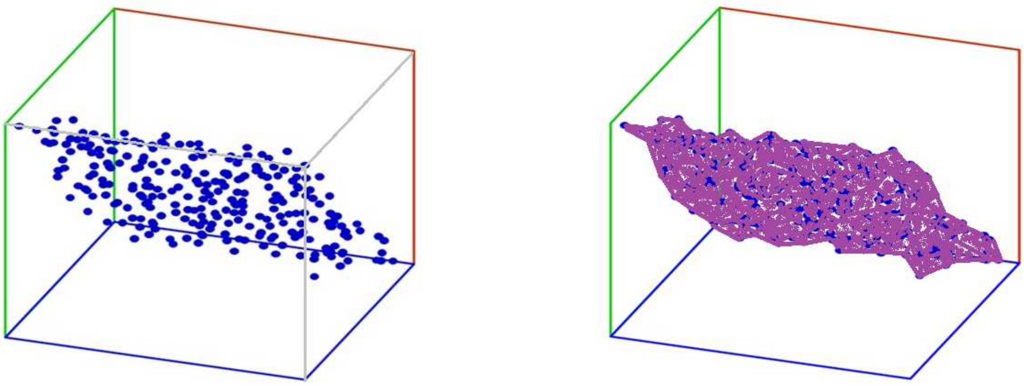
This plane is given by:

$$y - f(x_i, y_j) = \frac{\Delta f}{\Delta x}(S_{ij})(x - x_i) + \frac{\Delta f}{\Delta y}(T_{ij})(y - y_j)\,.$$

This plane restricted to the triangle formed by those three points is a *triangle* in our 3-space. For each (i, j), there four such triangles; just take $(i \pm 1, j \pm 1)$.

Exercise 3.7.7

Under what circumstances, when taken over all possible such pairs (i, j), do these triangles form a *mesh*? In other words, when do they fit together without breaks?



For a simplified notation, we will often *omit the indices*:

Difference quotients

$$\frac{\Delta f}{\Delta x}(s, y) = \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

$$\frac{\Delta f}{\Delta y}(x, t) = \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

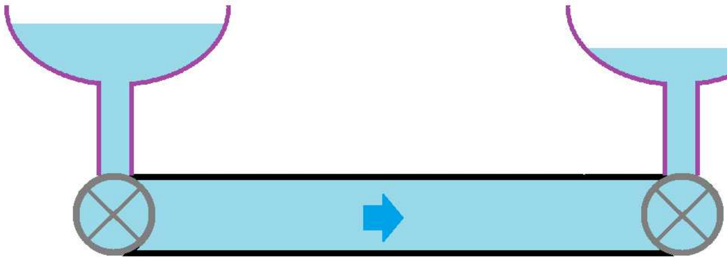
Warning!

The difference quotient is not a fraction.

Example 3.7.8: hydraulic analogy

A simple interpretation of this data is *water flow*, as follows:

- Each edge of the partition represents a pipe.
- The function $f(x_i, y_j)$ represents the water pressure at the joint (x_i, y_j).
- The difference of the pressure between any two adjacent joints causes the water to flow.
- The differences $\Delta_x f(s_i, y_j)$ and $\Delta_y f(x_i, t_j)$ are the flow amounts along these pipes.
- The difference quotients $\frac{\Delta f}{\Delta x}(s_i, y_j)$ and $\frac{\Delta f}{\Delta y}(x_i, t_j)$ are the flow rates along these pipes.

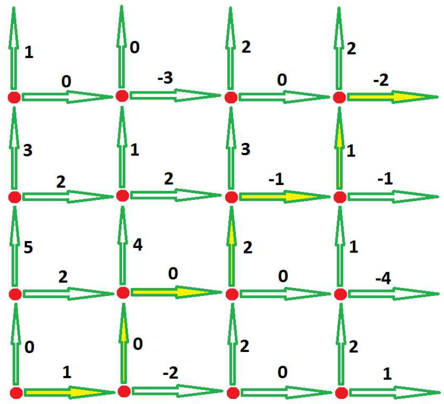


For *electric current*, substitute "electric potential" for "pressure".

When there are no secondary nodes specified, we can think of S_{ij} and T_{ij} as standing for the edges themselves:

- $S_{ij} = [x_i, x_{i+1}] \times \{y_j\}$, and
- $T_{ij} = \{x_i\} \times [y_j, y_{j+1}]$.

They are the inputs of the partial difference quotients. Taken as a whole, the difference may look like this:



We bring the two partial difference quotients together in the same way as the partial differences:

Definition 3.7.9: difference quotient

The *difference quotient* of a function of several variables defined at the primary nodes of a partition of a plane region is defined at secondary nodes by the following:

$$\boxed{\frac{\Delta f}{\Delta X}(N) = \begin{cases} \dfrac{\Delta_x f}{\Delta x}(S_{ij}) & \text{if } N = S_{ij}\,, \\ \dfrac{\Delta_y f}{\Delta y}(T_{ij}) & \text{if } N = T_{ij}\,. \end{cases}}$$

This is a real-valued 1-form.

Why aren't these vectors? From our study of planes, we know that the vector formed by these two numbers is especially important:

$$\boxed{\begin{cases} \left\langle \dfrac{\Delta_x f}{\Delta x}(S_{ij}), 0 \right\rangle & \text{if } N = S_{ij}\,, \\ \left\langle 0, \dfrac{\Delta_y f}{\Delta y}(T_{ij}) \right\rangle & \text{if } N = T_{ij}\,. \end{cases}}$$

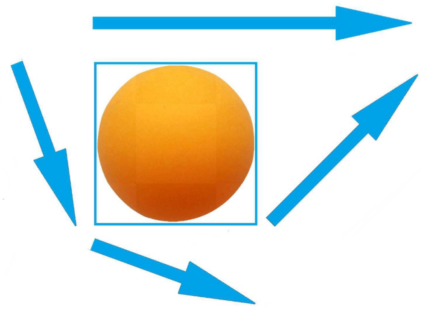
This is a vector-valued 1-form. There may be others, more complex.

Example 3.7.10: hydraulic analogy

According to the hydraulic analogy, only the flow *along* the pipe matters while any leakage is ignored. We can also see a flow on a surface. We consider *vector fields* (or vector-valued 1-forms); there is a vector assigned to each edge:

$$\begin{aligned} F(s,y) &= \langle \quad p(s,y) \quad , q(s,y) \quad \rangle, \\ F(x,t) &= \langle \quad r(s,y) \quad , s(x,t) \quad \rangle. \end{aligned}$$

In other words, there is also a component perpendicular to the corresponding edge:



However, only the *projections* of these vectors on the edges affect the rotation of the ball.

Definition 3.7.11: gradient vector field

A vector field F defined at each secondary node of the partition is called *gradient* if there is such a function $z = f(x,y)$ defined on the nodes of the partition that

$$\boxed{F(N) \cdot E = \Delta f\,(N)}$$

for every edge E and its secondary node N.

It follows that:

- The horizontal component of $F(N)$ is equal to $\frac{\Delta f}{\Delta x}(N)$ when N is on a horizontal edge.
- The vertical component of $F(N)$ is equal to $\frac{\Delta f}{\Delta y}(N)$ when N is a vertical edge.

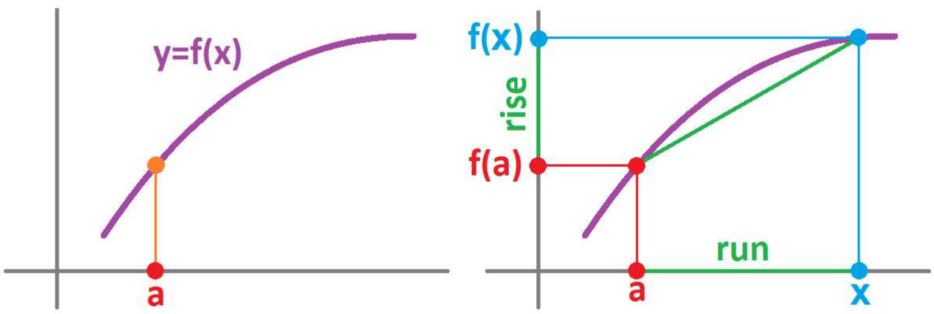
The other components are irrelevant.

Next, we continue in the same manner as before:

$$\lim_{\Delta X \to 0} \left(\begin{array}{c} \text{discrete} \\ \text{calculus} \end{array} \right) = \text{calculus}$$

3.8. *The average and the instantaneous rates of change*

Recall the 1-dimensional case. A linear approximation of a function $y = f(x)$ at $x = a$ is function that defines a secant line, i.e., a line on the xy-plane through the point of interest $(a, f(a))$ and another point on the graph $(x, f(x))$:

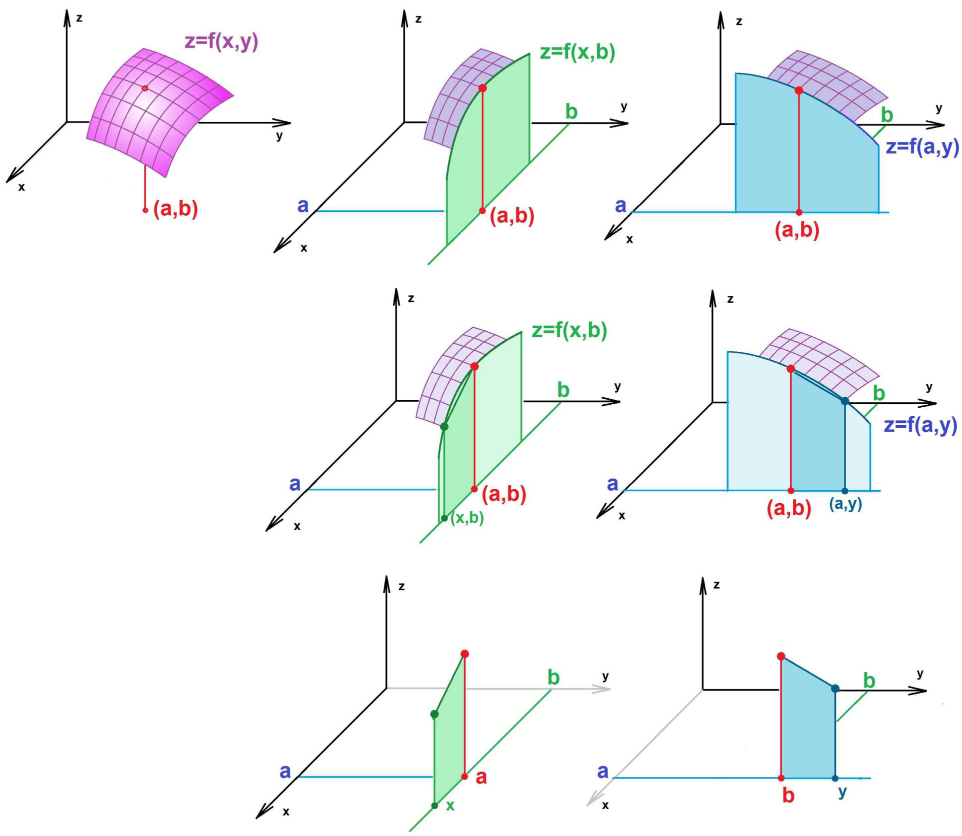


Its slope is the difference quotient of the function:

$$\frac{\Delta f}{\Delta x} = \frac{f(x) - f(a)}{x - a} .$$

Now, let's see how this plan applies to functions of two variables.

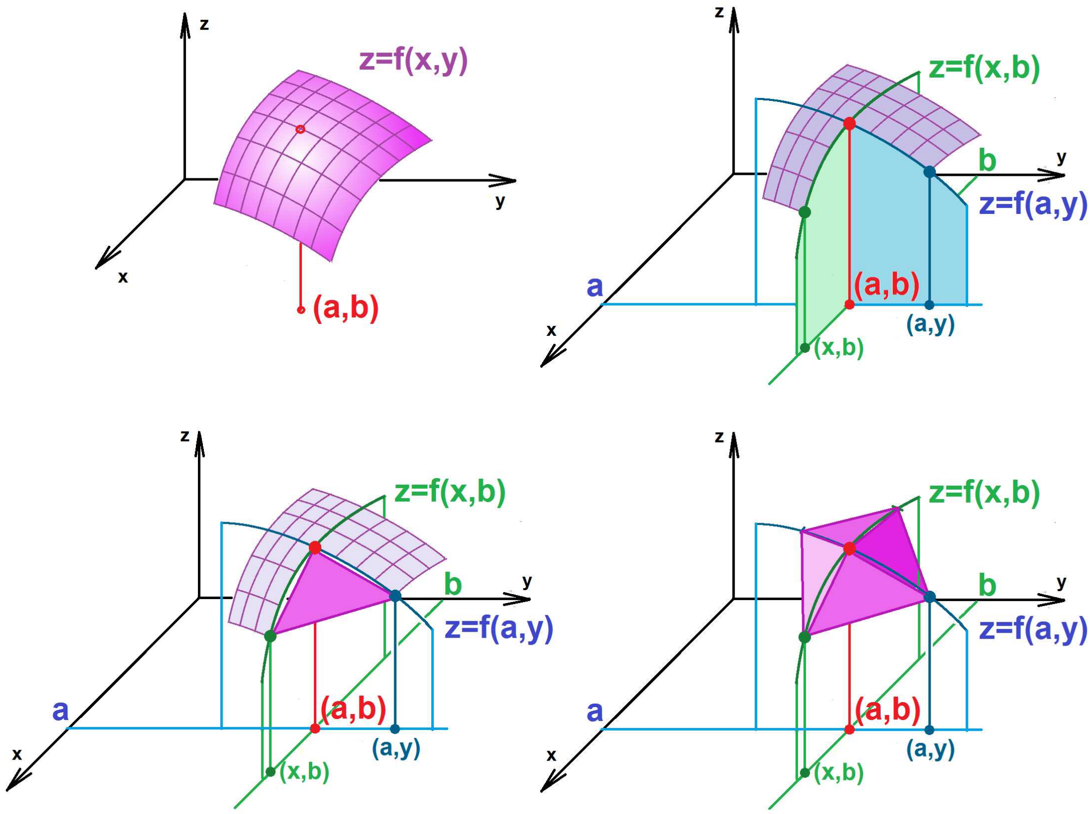
A linear approximation of a function $z = f(x, y)$ at $(x, y) = (a, b)$ is a function that represents a secant *plane*, i.e., a plane in the xyz-space through the point of interest $(a, b, f(a, b))$ and *two* other points on the graph. In order to ensure that these points define a plane, they should be chosen in such a way that they aren't on the same line. The easiest way to accomplish that is to choose the last two to lie in the x- and the y-directions from (a, b), i.e., (x, b) and (a, y) with $x \neq a$ and $y \neq b$.



The two slopes in these two directions are the two difference quotients, with respect to x and with respect to y:

$$\frac{\Delta f}{\Delta x} = \frac{f(x, b) - f(a, b)}{x - a} \quad \text{and} \quad \frac{\Delta f}{\Delta y} = \frac{f(a, y) - f(a, b)}{y - b} .$$

The two lines form the *secant plane*.



Definition 3.8.1: partial derivatives

The *partial derivatives of f with respect to x and y at $(x,y)=(a,b)$* are defined to be the limits of the partial difference quotients with respect to x at $x=a$ and with respect to y at $y=b$ respectively, denoted as follows:

$$\boxed{\frac{\partial f}{\partial x}(a,b)=\lim_{x\to a}\frac{f(x,b)-f(a,b)}{x-a}\quad\text{and}\quad\frac{\partial f}{\partial y}(a,b)=\lim_{y\to b}\frac{f(a,y)-f(a,b)}{y-b}}$$

as well as

$$\boxed{f'_x(a,b)\quad\text{and}\quad f'_y(a,b)}$$

The following is an obvious conclusion.

Theorem 3.8.2: Partial Derivatives

The partial derivatives of f at $(x,y)=(a,b)$ are found as the derivatives of f with respect to x and to y:

$$\boxed{\frac{\partial f}{\partial x}(a,b)=\frac{d}{dx}f(x,b)\bigg|_{x=a}\quad\text{and}\quad\frac{\partial f}{\partial y}(a,b)=\frac{d}{dy}f(a,y)\bigg|_{y=b}}$$

As a result, the computations are straight-forward.

Example 3.8.3: a computation

Find the partial derivatives of the function:

$$f(x,y)=(x-y)e^{x+y^2}$$

at $(x,y)=(0,0)$:

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,0) &= \frac{d}{dx}(x-0)e^{x+0^2}\Big|_{x=0} &= \frac{d}{dx}xe^{x}\Big|_{x=0} &= e^x+xe^x\Big|_{x=0} &= 1\,,\\
\frac{\partial f}{\partial y}(0,0) &= \frac{d}{dy}(0-y)e^{0+y^2}\Big|_{y=0} &= \frac{d}{dy}-ye^{y^2}\Big|_{y=0} &= -e^{y^2}-ye^{y^2}2y &= -e^{y^2}-2y^2e^{y^2}\,.
\end{aligned}$$

Exercise 3.8.4

Find the partial derivatives of the function:

$$f(x,y)=?$$

Exercise 3.8.5

Now in reverse. Find a function f of two variables the partial derivatives of which are these:

$$\frac{\partial f}{\partial x}=?\quad \frac{\partial f}{\partial y}=?$$

3.9. *Linear approximations and differentiability*

This is the build-up for the introduction of the derivative of functions of two variables:

- the slopes of a surface in different directions
- the secant lines
- the secant planes
- the planes and ways to represent them
- the limits

Due to the complexity of the problem we will focus on the *derivative at a single point* for now.

Let's review what we know about the derivatives so far. The definition of the derivative of a parametric curve (at a point)is virtually identical to that for a numerical function because the fraction of the difference quotient is allowed by vector algebra:

numerical functions

$$f'(a)=\lim_{x\to a}\frac{f(x)-f(a)}{x-a}$$

parametric curves

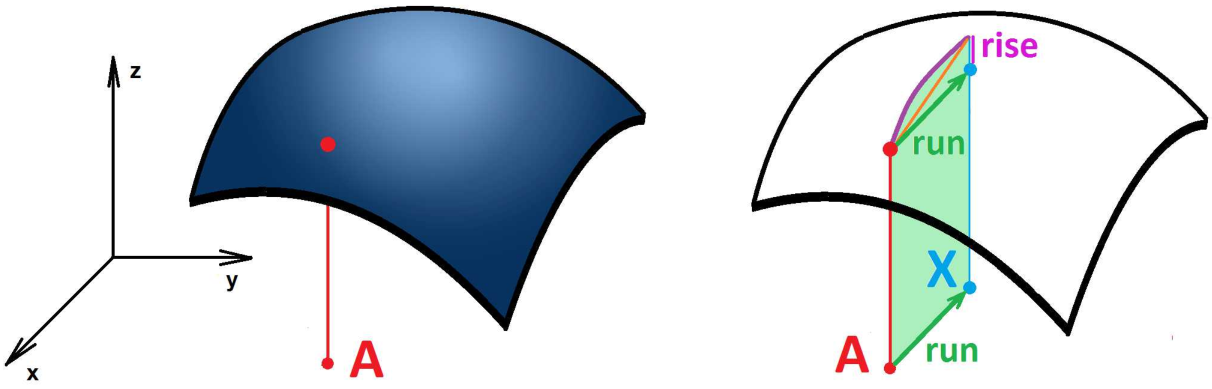
$$F'(a)=\lim_{t\to a}\frac{F(t)-F(a)}{t-a}$$

functions of several variables

$$f'(A)=\lim_{X\to A}\frac{f(X)-f(A)}{X-A}\ ???$$

The same formula fails for a function of several variables. We can't divide by a vector! This failure is the reason why we start studying multi-dimensional calculus with parametric curves and not functions of several variables.

Can we fix the definition?



How about we divide by the *magnitude* of this vector? This is allowed by vector algebra and the result is something similar to the rise over the run definition of the slope.

Example 3.9.1: linear function

Let's carry out this idea for a linear function, the simplest kind of function:

$$f(x,y)=2x+y\,.$$

The limit:

$$\lim_{(x,y)\to(0,0)}\frac{|2x+y|}{\sqrt{x^2+y^2}}\,,$$

is to be evaluated along either of the axes:

$$\lim_{x\to0}\frac{|2x+0|}{\sqrt{x^2+0^2}}=\lim_{x\to0}\frac{|2x|}{|x|}=2\neq\lim_{y\to0}\frac{|2\cdot0+y|}{\sqrt{0^2+y^2}}=\lim_{y\to0}\frac{|y|}{|y|}=1\,.$$

The limit doesn't exist because the slopes are different in different directions.

The line of attack of the meaning of the derivative then shifts – to finding that tangent plane. Some of the functions shown in the last section have no tangent planes at their points of discontinuity. But such a plane is just a linear function! As such, it is a linear approximation of the function.

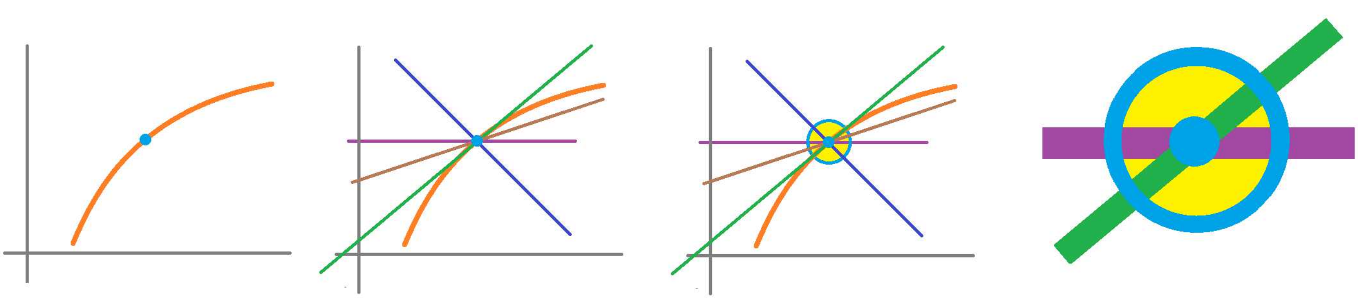
In the 1-dimensional case, among the linear approximations of a function $y=f(x)$ at $x=a$ are the functions that define secant lines, i.e., lines on the xy-plane through the point of interest $(a,f(a))$ and another point on the graph $(x,f(x))$. Its slope is the difference quotient of the function:

$$\frac{\Delta f}{\Delta x}=\frac{f(x)-f(a)}{x-a}\,.$$

In general, they are just linear functions and their graphs are lines through $(a,f(a))$; in the point-slope form they are:

$$l(x)=f(a)+m(x-a).$$

When you zoom in on the point, the tangent line – but no other line – will merge with the graph:



We reformulate our theory from Volume 2 (Chapter 2DC-6) slightly.

Definition 3.9.2: best linear approximation

Suppose $y = f(x)$ is a defined at $x = a$ and

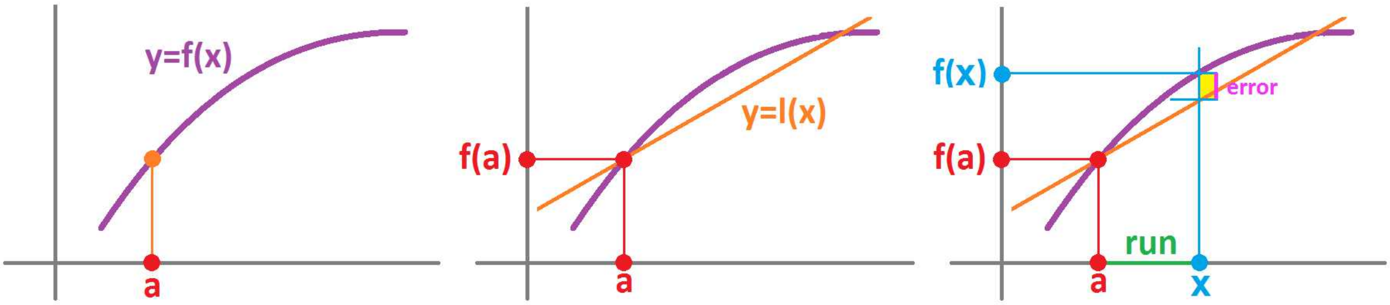
$$l(x) = f(a) + m(x - a)$$

is any of its linear approximations at that point. Then, $y = l(x)$ is called the *best linear approximation* of f at $x = a$ if the following is satisfied:

$$\lim_{x \to a} \frac{f(x) - l(x)}{|x - a|} = 0\,.$$

The definition is about the decline of the error of the approximation, i.e., the difference between the two functions:

$$\text{error } = |f(x) - l(x)|.$$



This, not only the error vanishes, but also it vanishes relative to how close we are to the point of interest, i.e., the run. The following result is from Volume 2 (Chapter 2DC-6):

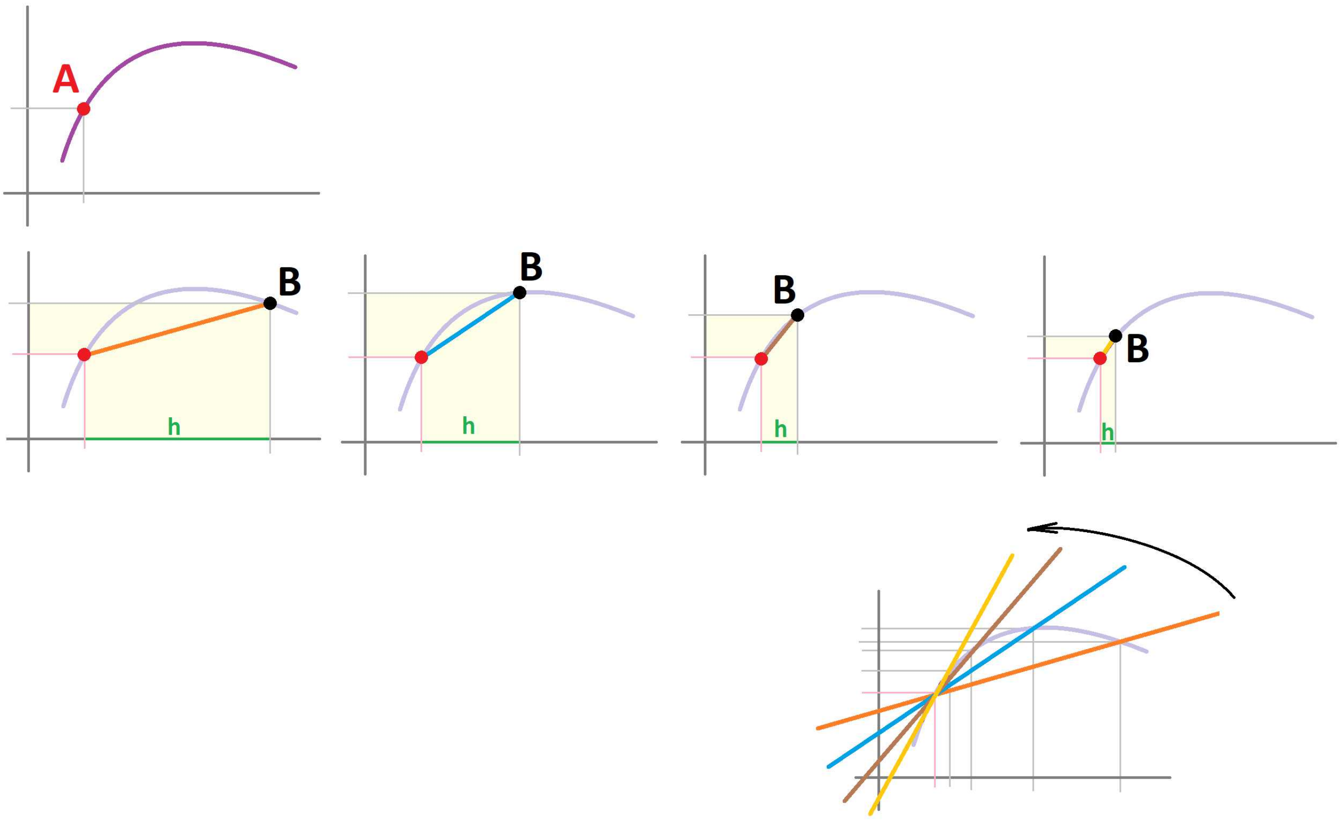
Theorem 3.9.3: Best Linear Approximation For One Variable

If

$$l(x) = f(a) + m(x - a)$$

is the best linear approximation of f at $x = a$, then

$$m = f'(a)\,.$$



Therefore, the graph of this function is the tangent line.

Now the partial derivatives of f with respect to x and y at $(x, y) = (a, b)$ are:

$$\frac{\partial f}{\partial x}(a,b) = \lim_{x \to a} \frac{f(x,b) - f(a,b)}{x - a} \quad \text{and} \quad \frac{\partial f}{\partial y}(a,b) = \lim_{y \to b} \frac{f(a,y) - f(a,b)}{y - b}.$$

From our study of planes, we know that the vector formed by these functions that is especially important:

Definition 3.9.4: gradient

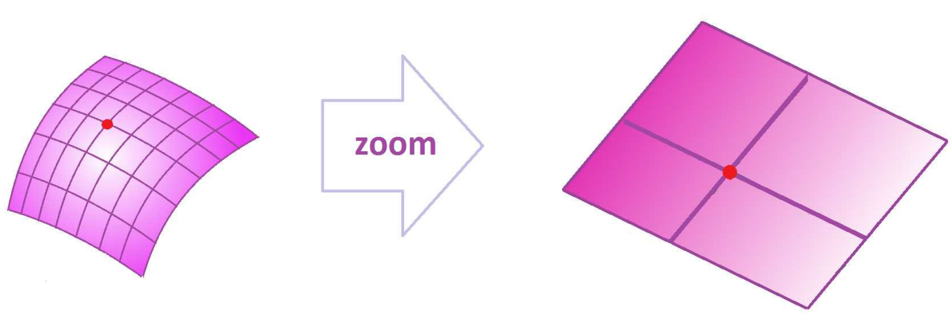
The *gradient* of f at $(x, y) = (a, b)$ is defined to be the limit of the difference quotient as well as the vector of partial derivatives denoted as follows:

$$\boxed{\nabla f(a,b) = \frac{df}{dX}(a,b) = \left\langle \frac{\partial f}{\partial x}(a,b), \frac{\partial f}{\partial y}(a,b) \right\rangle}$$

In general these approximations are just linear functions and their graphs are planes through the point $(a, b, f(a, b))$; in the point-slope form they are:

$$l(x,y) = f(a,b) + m(x-a) + n(y-b).$$

When you zoom in on the point, the tangent plane – but no other plane – will merge with the graph:



Definition 3.9.5: best linear approximation

Suppose $y = f(x, y)$ is defined at $(x, y) = (a, b)$ function and

$$l(x, y) = f(a, b) + m(x - a) + n(y - b)$$

is any of its linear approximations at that point. Then, $y = l(x, y)$ is called the *best linear approximation* of f at $(x, y) = (a, b)$ if the following is satisfied:

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - l(x, y)}{||(x, y) - (a, b)||} = 0\,.$$

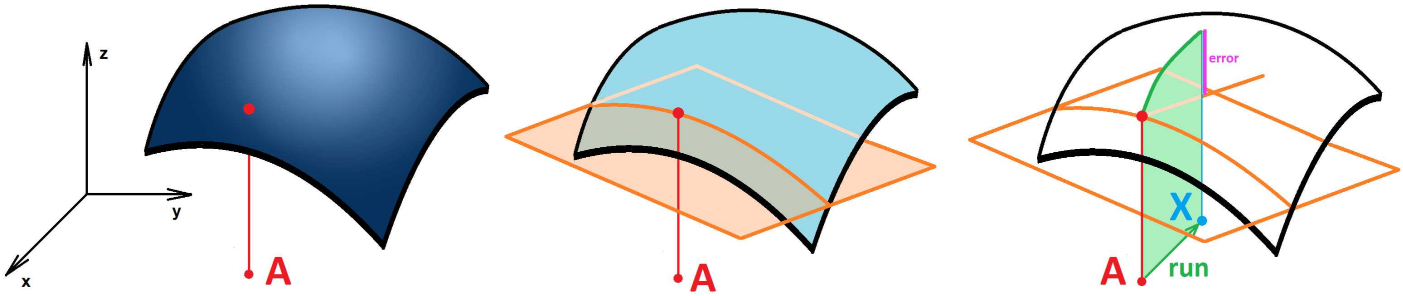
In that case, the function f is called *differentiable* at (a, b) and the graph of $z = l(x, y)$ is called the *tangent plane*.

In other words, we stick to the functions that look like plane on a small scale!

The definition is about the decline of the error of the approximation:

$$\text{error } = |f(x, y) - l(x, y)|\,.$$

The limit of $\dfrac{\text{error}}{\text{run}}$ is required to be zero.



Theorem 3.9.6: Best Linear Approximation For Two Variables

If

$$l(x, y) = f(a, b) + m(x - a) + n(y - b)$$

is the best linear approximation of f at $x = a$, then

$$m = \frac{\partial f}{\partial x}(a, b) \quad \text{and} \quad n = \frac{\partial f}{\partial y}(a, b)\,.$$

Proof.

The limit in the definition allows us to approach (a, b) in any way we like. Let's start with the direction parallel to the x-axis and see how $\dfrac{\text{error}}{\text{run}}$ in converted into $\dfrac{\text{rise}}{\text{run}}$:

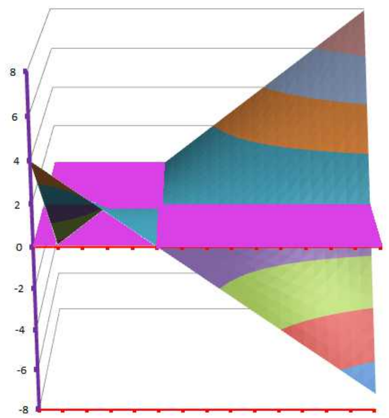
$$\begin{aligned}
0 \ &= \lim_{x\to a^+,\ y=b} \frac{f(x, y) - l(x, y)}{||(x, y) - (a, b)||} \\
&= \lim_{x\to a^+} \frac{f(x, y) - (f(a, b) + m(x - a) + n(b - b))}{x - a} \\
&= \lim_{x\to a^+} \frac{f(x, y) - f(a, b)}{x - a} - m \\
&= \frac{\partial f}{\partial x}(a, b) - m\,.
\end{aligned}$$

The same computation for the y-axis produces the second identity.

Example 3.9.7: hyperbolic paraboloid

Let's confirm that the tangent plane to the hyperbolic paraboloid, given by the graph of $f(x,y)=xy$, at $(0,0)$ is horizontal:

$$\frac{\partial f}{\partial x}(0,0)=\frac{d}{dx}f(x,0)\Big|_{x=0}=\frac{d}{dx}(x\cdot 0)\Big|_{x=0}=0,\ \frac{\partial f}{\partial y}(0,0)=\frac{d}{dy}f(0,y)\Big|_{y=0}=\frac{d}{dx}(0\cdot y)\Big|_{y=0}=0\,.$$



The two tangent lines form a cross and the tangent plane is spanned on this cross.

Just as before, replacing a function with its linear approximation is called *linearization* and it can be used to estimate values of new functions.

Example 3.9.8: approximation

Let's review a familiar example: approximation of $\sqrt{4.1}$. We can't compute $\sqrt{x}$ by hand because, in a sense, the function $f(x)=\sqrt{x}$ is *unknown*. The best linear approximation of $y=f(x)=\sqrt{x}$ is known and, as a linear function, it can be computed by hand:

$$f'(x)=\frac{1}{2\sqrt{x}}\implies f'(4)=\frac{1}{2\sqrt{4}}=\frac{1}{4}\,.$$

The best linear approximation is:

$$l(x)=f(a)+f'(a)(x-a)=2+\frac{1}{4}(x-4)\,.$$

Finally, our approximation of $\sqrt{4.1}$ is

$$l(4.1)=2+\frac{1}{4}(4.1-4)=2+\frac{1}{4}\cdot 1=2+0.025=2.025\,.$$

Let's now approximate of $\sqrt{4.1}\sqrt[3]{7.8}$. Instead of approximating the two terms separately, we will find the best linear approximation of the product $f(x,y)=\sqrt{x}\sqrt[3]{y}$ at $(x,y)=(4,8)$. Then we compute the partial derivatives, x:

$$\frac{\partial f}{\partial x}(4,8)=\frac{d}{dx}f(x,8)\Big|_{x=4}=\frac{d}{dx}\left(\sqrt{x}\sqrt[3]{8}\right)\Big|_{x=4}=\frac{1}{2\sqrt{x}}2\Big|_{x=4}=\frac{1}{2}\,.$$

And y:

$$\frac{\partial f}{\partial y}(4,8)=\frac{d}{dy}f(4,y)\Big|_{y=8}=\frac{d}{dx}\left(\sqrt{4}\sqrt[3]{y}\right)\Big|_{y=8}=2\frac{1/3}{\sqrt[3]{y^2}}\Big|_{y=8}=\frac{1}{12}\,.$$

The best linear approximation is:

$$l(x,y)=f(a,b)+\frac{\partial f}{\partial x}(a,b)(x-a)+\frac{\partial f}{\partial y}(a,b)(y-b)=2\cdot 2+\frac{1}{2}(x-4)+\frac{1}{12}(y-8)\,.$$

Finally, our approximation of $\sqrt{4.1}\sqrt[3]{7.8}$ is

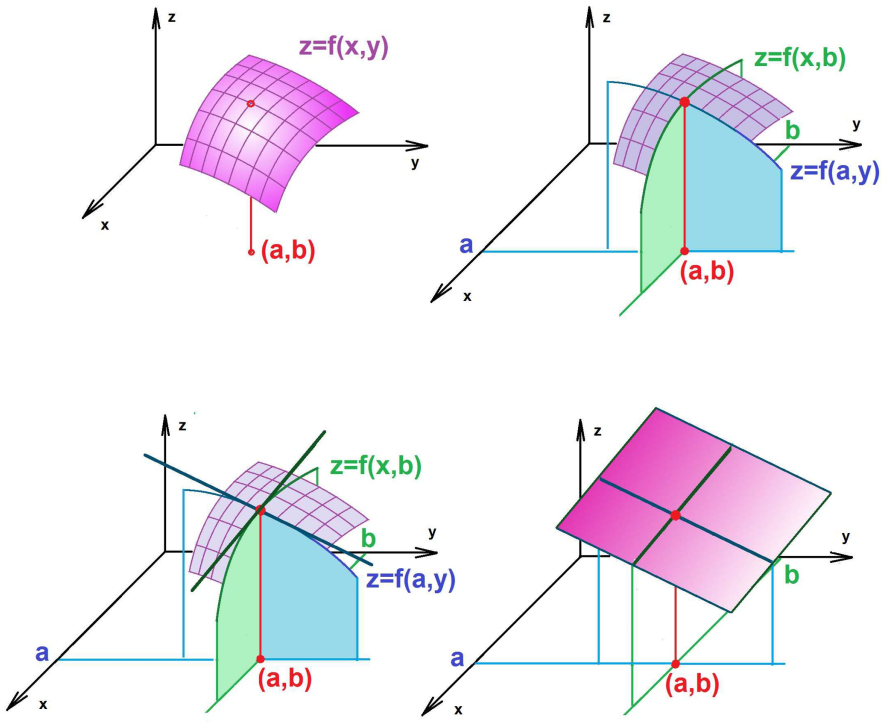
$$l(4.1,7.8)=4+\frac{1}{2}.1+\frac{1}{12}(-.2)=4.033333...$$

Exercise 3.9.9

Find the best linear approximation of $f(x)=(xy)^{1/3}$ at $(1,1)$.

To summarize:

- Under the limit $x \to a$, the secant line parallel to the x-axis turns into a tangent line with the slope $f'_x(a,b)$.
- Under the limit $y \to b$, the secant line parallel to the x-axis turns into a tangent lines with the slope $f'_y(a,b)$, provided f is differentiable with respect to these two variables at this point. Meanwhile,
- The secant plane turns into the tangent plane, provided f is differentiable at this point, with the slopes $f'_x(a,b)$ and $f'_y(a,b)$ in the directions of the coordinate axes.



3.10. *Partial differentiation and optimization*

We pick one variable and treat the rest of them as parameters.

The rules are exactly the same.

This is, for example, the *Constant Multiple Rule*:

$$\frac{\partial}{\partial x}(xy)=y\frac{\partial}{\partial x}(x)=y\cdot 1=y\,,\ \frac{\partial}{\partial y}(xy)=x\frac{\partial}{\partial y}(y)=x\cdot 1=x\,.$$

This is the *Sum Rule*:

$$\frac{\partial}{\partial x}(x+y)=\frac{\partial}{\partial x}(x)+\frac{\partial}{\partial x}(y)=1+0=1\,,\ \frac{\partial}{\partial y}(x+y)=\frac{\partial}{\partial y}(x)+\frac{\partial}{\partial y}(y)=0+1=1\,.$$

Example 3.10.1: related rates?

This notion is not to be confused with the "related rates" (Volume 2). Here y is a function of x:

$$\frac{d}{dx}(xy^2) = y^2 + x \cdot \frac{d}{dx}(y^2) = y^2 + x \cdot 2yy',$$

by the *Chain Rule*. Here y is just another variable:

$$\frac{\partial}{\partial x}(xy^2) = y^2 + x \cdot \frac{\partial}{\partial x}(y^2) = y^2 + x \cdot 0 = y^2.$$

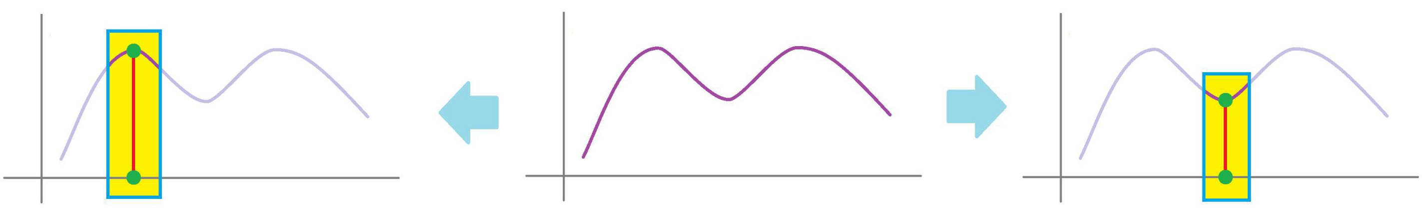
After all, these variable are *un*-related; they are *independent* variables!

Next we consider *optimization*, which is simply a way to find the maximum and minimum values of functions.

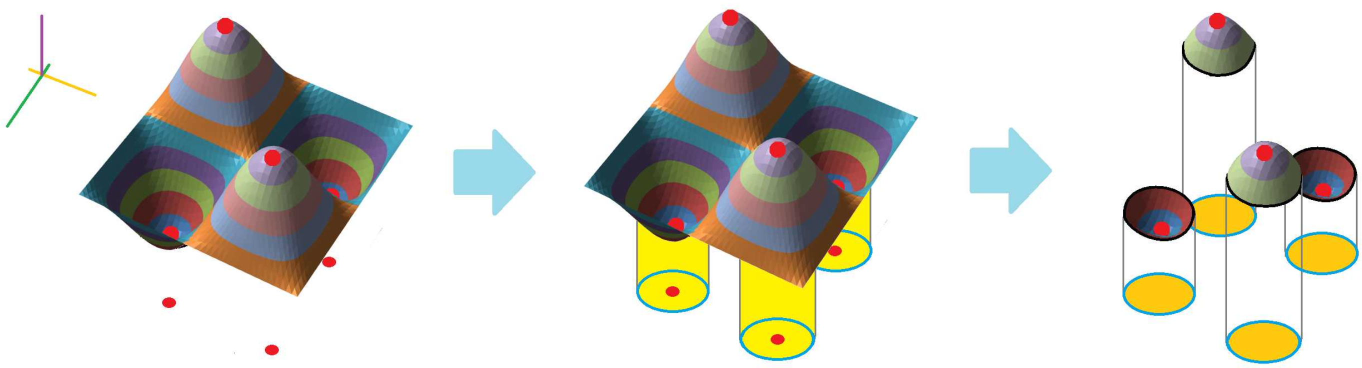
We already know from the *Extreme Value Theorem* that this problem always has a solution provided the function is continuous on a closed bounded subset.

Furthermore, just as in Volume 2 (Chapter 2DC-5), we first narrow down our search.

Recall that a function $y = f(x)$ has a local minimum point at $x = a$ if $f(a) \leq f(x)$ for all x within some open interval centered at a. We can imagine as if we build a rectangle on top of each of these intervals and use to cut a piece from our graph:



We can't use intervals in dimension 2, what's their analog? It is an open disk on the plane centered at (a, b). We can imagine as if we build a cylinder on top of each of these disks and use to cut a patch from the surface of our graph:



In fact, both intervals and disks (and 3d balls, etc.) can be conveniently described in terms of the distance from the point: $d(X, A) \leq \varepsilon$. So, we restrict our attention to these (possibly small) disks.

Definition 3.10.2: local maximum and minimum points

A function $z = f(X)$ has a *local minimum point* at $X = A$ if $f(A) \leq f(X)$ for all X within some positive distance from A, i.e., $d(X, A) \leq \varepsilon$. Furthermore, a function f has a *local maximum point* at $X = A$ if $f(A) \geq f(X)$ for all X for all X within some positive distance from A. We call these *local extreme points*, or extrema.

Warning!

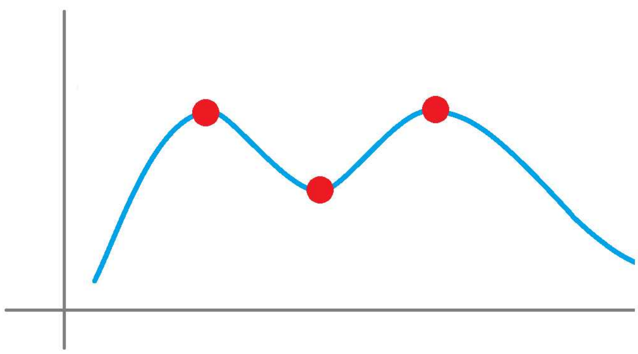
The definition implies that f is defined for all of these values of X.

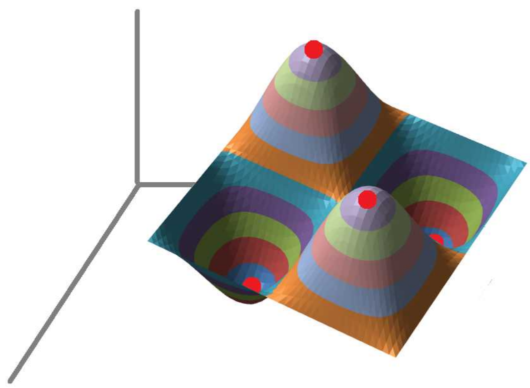
Exercise 3.10.3

What could be a possible meaning of an analog of the one-sided derivative for functions of two variables?

In other words, there is an open disk D around A such that A is the global maximum (or minimum) point when f is restricted to A.

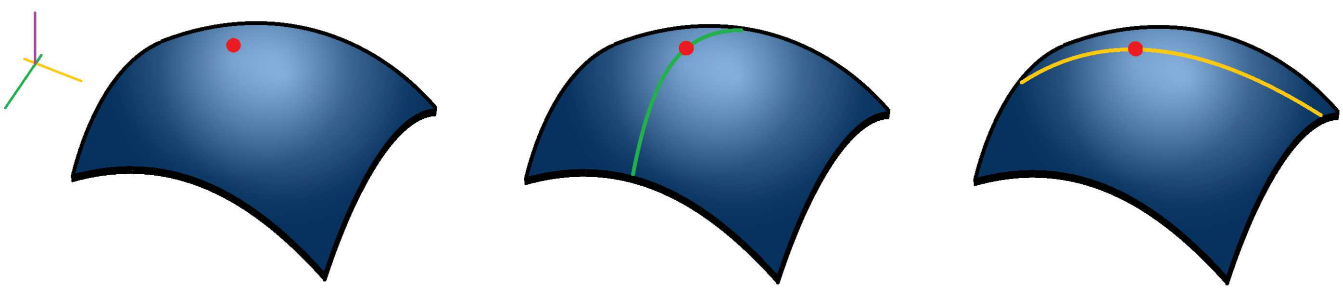
Now, local extreme points are *candidates* for global extreme points. To compare to the familiar, one-dimensional case, we use to look at the terrain *from the side* and now we look from above:





How do we find them? We go from the two-dimensional problem to the one-dimensional case. Indeed, if (a, b) is a local maximum of $z = f(x, y)$ then

- $x = a$ is a local maximum of the numerical function $g(x) = f(x, b)$, and
- $y = b$ is a local maximum of the numerical function $h(y) = f(a, y)$.



In other words, the summit will be the highest point of our trip whether we come from the south or from the west.

Warning!

with this approach we ignore the possibility of even higher locations found in the diagonal direction. However, the danger of missing this is only possible when the function isn't differentiable.

Just as in Volume 2 (Chapter 2DC-5), we will use the derivatives in order to facilitate our search. Recall what *Fermat's Theorem* states: if

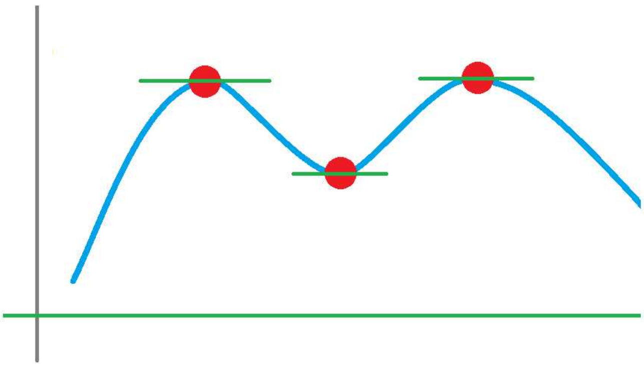
- $x = a$ is a local extreme point of $z = f(x)$ and
- $z = f(x)$ is differentiable at $x = a$,

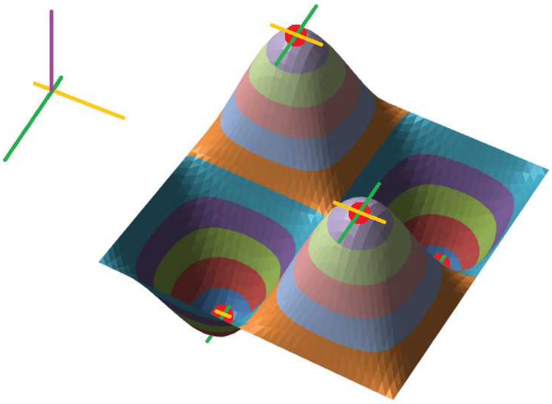
then a is a *critical point*, i.e.,

$$f'(a) = 0\,,$$

or undefined.

The last equation typically produces just a finite number of *candidates* for extrema. As we apply the theorem to either variable, we find ourselves in a similar situation but with *two* equations.





For a function of two variables we have the following two-dimensional analog of Fermat's Theorem.

Theorem 3.10.4: Fermat's Theorem

If $X = A$ is a local extreme point of a differentiable at A function several variables $z = f(X)$ then A is a critical point, i.e.,

$$\boxed{\nabla f\,(A) = 0}$$

Example 3.10.5: critical points

Let's find the critical points of

$$f(x, y) = x^2 + 3y^2 + xy + 7\,.$$

We differentiate:

$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) &= 2x + y\,,\\ \frac{\partial f}{\partial x}(x, y) &= 6y + x\,.\end{aligned}$$

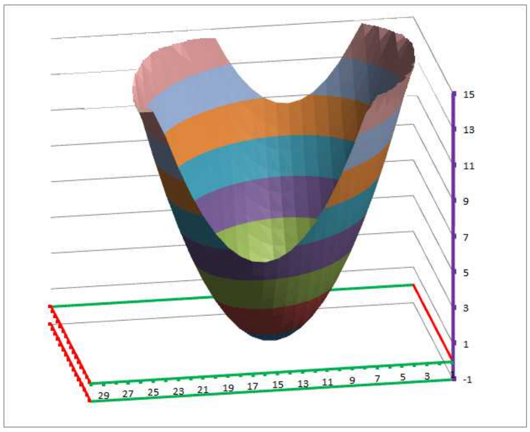
We how set these derivatives equal to zero:

$$2x + y = 0,$$
$$6y + x = 0.$$

We have two equations to be solved. Geometrically, these are two lines and the solution is the intersection. By substitution:

$$6y + x = 0 \implies x = -6y \implies 2(-6y) + y = 0 \implies y = 0 \implies x = 0\,.$$

There is only one critical point $(0, 0)$. It is just a candidate for an extreme point, for now. Plotting reveals that this is a minimum. This is what the graph looks like:



This is a *elliptic paraboloid* (its level curves are ellipses).

Instead of one equation - one variable as in the case of a numerical function, we will have two variables - two equations (and then three variables - three equations, etc.). Typically the result is finitely many points.

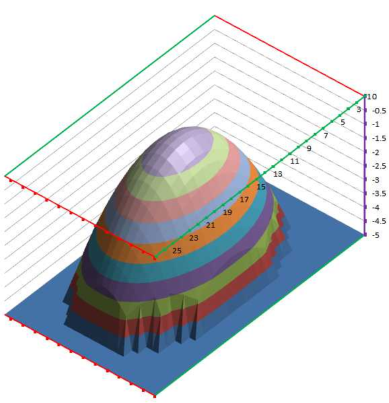
Exercise 3.10.6

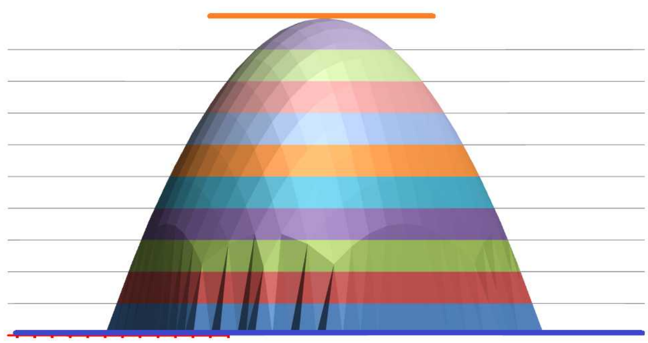
Give an example of a differentiable function of two variables with an infinitely many critical points.

Of course, the condition of the theorem – partial derivatives are zero – can be re-stated in a vector form – the gradient is zero:

$$f_x(a,b) = 0 \text{ and } f_y(a,b) = 0 \iff \nabla f(a,b) = 0\,.$$

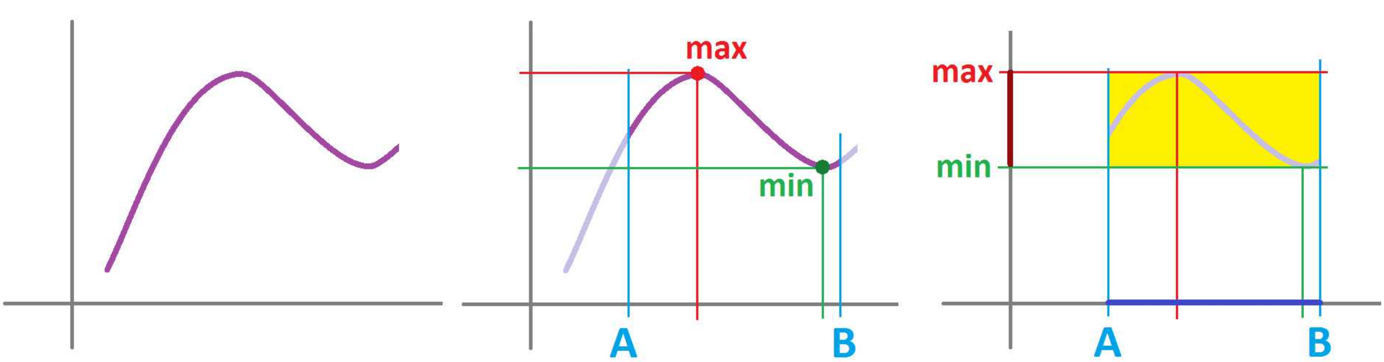
This means that the *tangent plane is horizontal*. So, at the top of the mountain, if it's smooth enough, it resembles a plane and this plane cannot be tilted because if it slopes down in any direction then so does the surface:





What we have seen is *local* optimization and it is only a stepping stone for *global* optimization.

Just as in the one-dimensional case, the *boundary* points have to be dealt with separately. This time, however, there will be infinitely many of them instead of just two end-points:



Example 3.10.7: extreme points

Let's find the extreme points of

$$f(x,y)=x^2+y^2\,,$$

subject to the restriction (of the domain):

$$|x|\le 1,\ |y|\le 1\,.$$

The restriction is the restriction of the domain of the function to this square.

We differentiate, set the derivatives equal to zero, and solve the system of equations:

$$\left.\begin{array}{lllll} f_x(x,y) & =2x & =0 & \Longrightarrow & x=0\\ f_y(x,y) & =2y & =0 & \Longrightarrow & y=0 \end{array}\right. \Longrightarrow\ (a,b)=(0,0)$$

This is the only critical point. We then note that

$$f(0,0)=0$$

before proceeding.

The boundary points of our square domain remain to be investigated:

$$|x|=1,\ |y|=1\,.$$

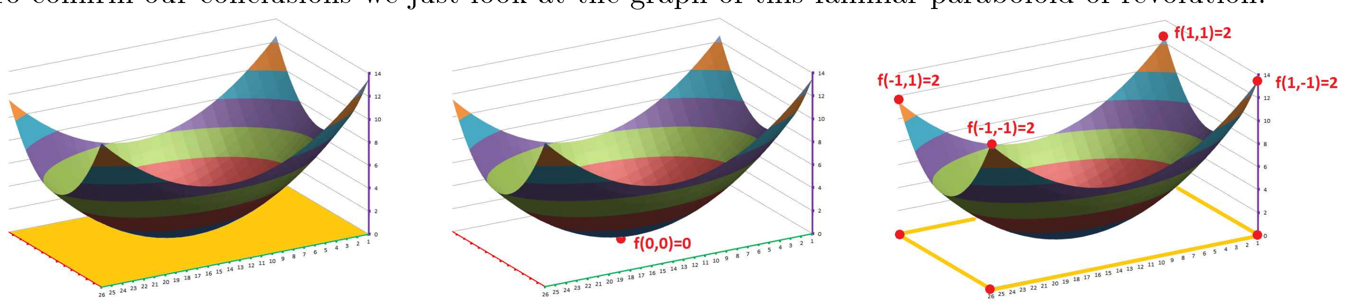
We have to test all of them by comparing the value of the function to each other and to the critical point. In other words, we are solving several one-dimensional optimization problems! These are the results:

$$\begin{array}{lllll} |x|=1 & \Longrightarrow & f(x,y)=1+y^2 & \Longrightarrow & \max\limits_{|y|\le 1}\{1+y^2\}=2 \text{ for } y=\pm1, \quad \min\limits_{|y|\le 1}\{1+y^2\}=1 \text{ for } y=0\\ |y|=1 & \Longrightarrow & f(x,y)=x^2+1 & \Longrightarrow & \max\limits_{|x|\le 1}\{x^2+1\}=2 \text{ for } x=\pm1, \quad \min\limits_{|x|\le 1}\{x^2+1\}=1 \text{ for } x=0 \end{array}$$

Comparing these outputs we conclude that:

- The maximal value of 2 is attained at the four points $(1,1),(-1,1),(1,-1),(-1,-1)$.
- The minimal value of 0 is attained at $(0,0)$.

To confirm our conclusions we just look at the graph of this familiar paraboloid of revolution:



Under the domain restrictions, the graph doesn't look like a cup anymore... We can see the vertical parabolas we just maximized.

For functions of three variables, the domains are typically solids. The boundaries of these domains are then surfaces! Then, a part of such a 3d optimization problem will be a 2d optimization problem...

We can think of the pair of partial derivatives as one, *the* derivative, of a function of two variables. However, in sharp contrast to the one-variable case, the resulting function has a very different nature from the original: same input but the output isn't a number anymore but a vector! It's a *vector field* discussed in Chapter 4.

3.11. The second difference quotient with respect to a repeated variable

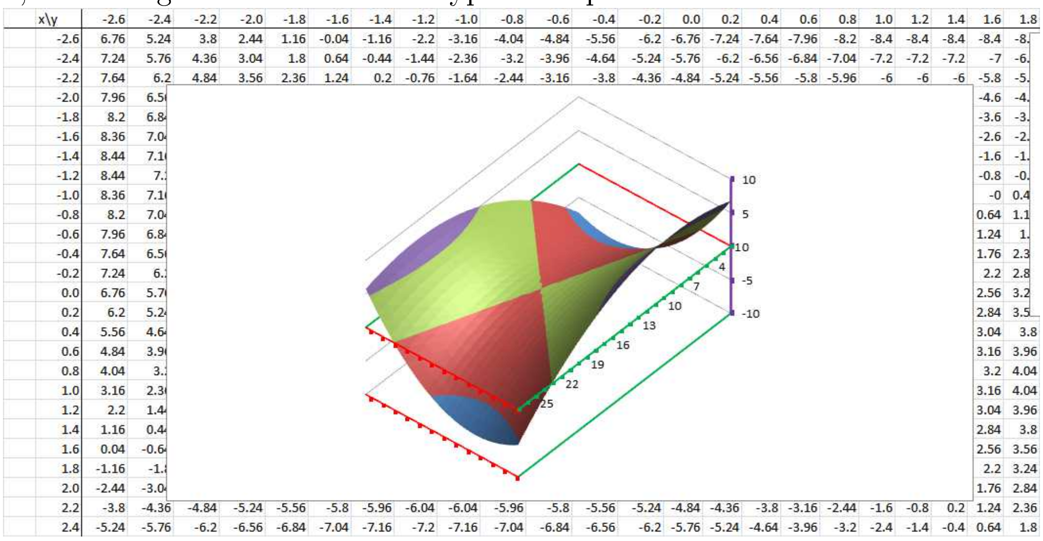
We start with *numerical* differentiation.

Example 3.11.1: a computation

We consider the function:

$$f(x,y) = -x^2 + y^2 + xy\,.$$

When plotted, it is recognized as a familiar hyperbolic paraboloid.



Below we outline the process of *second partial differentiation*. The variable functions – with respect to x and y – are shown and so are their derivatives. These are the functions we will differentiate.

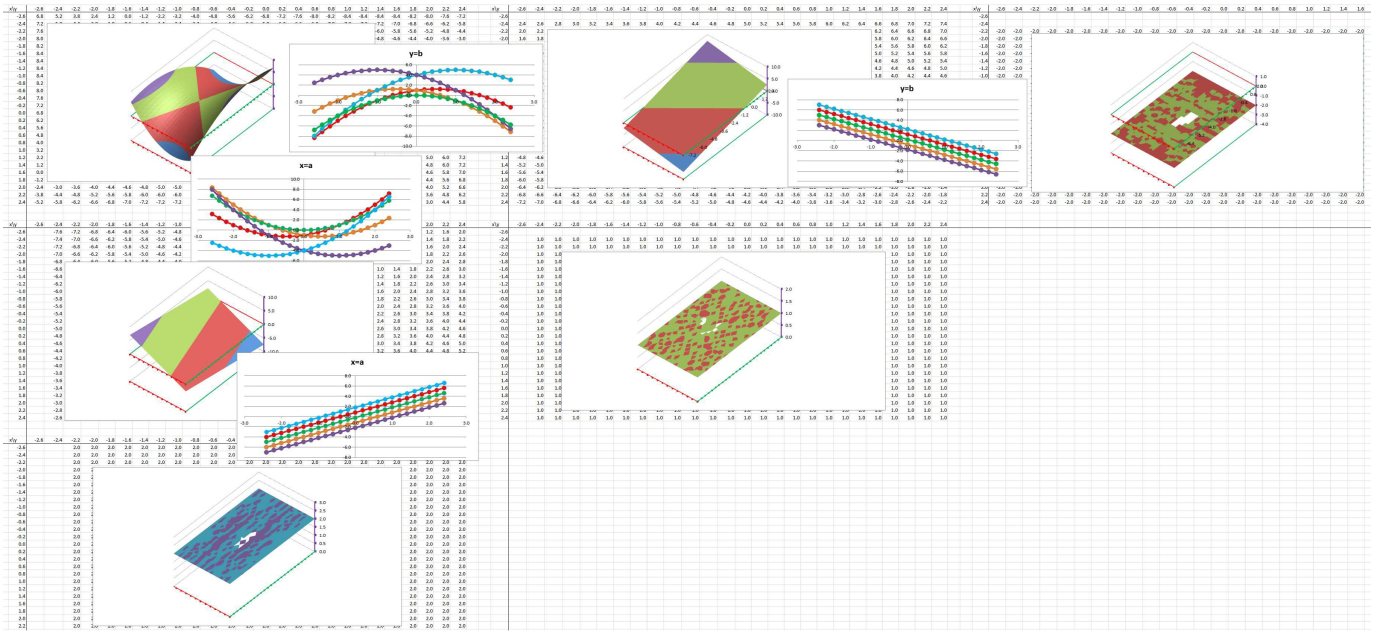
First, for each value of y given in the top row, we compute the difference quotient function with respect to x by going down the corresponding column and then placing these values on right in a new table (it is one row short in comparison to the original):

```
=(RC[-29]-R[-1]C[-29])/R2C1
```

Second, for each value of x given in the left-most column, we compute the difference quotient function with respect to y by going right the corresponding row and then placing these values below in a new table (it is one column short in comparison to the original).

```
=(R[-29]C-R[-29]C[-1])/R2C1
```

This is the summary:

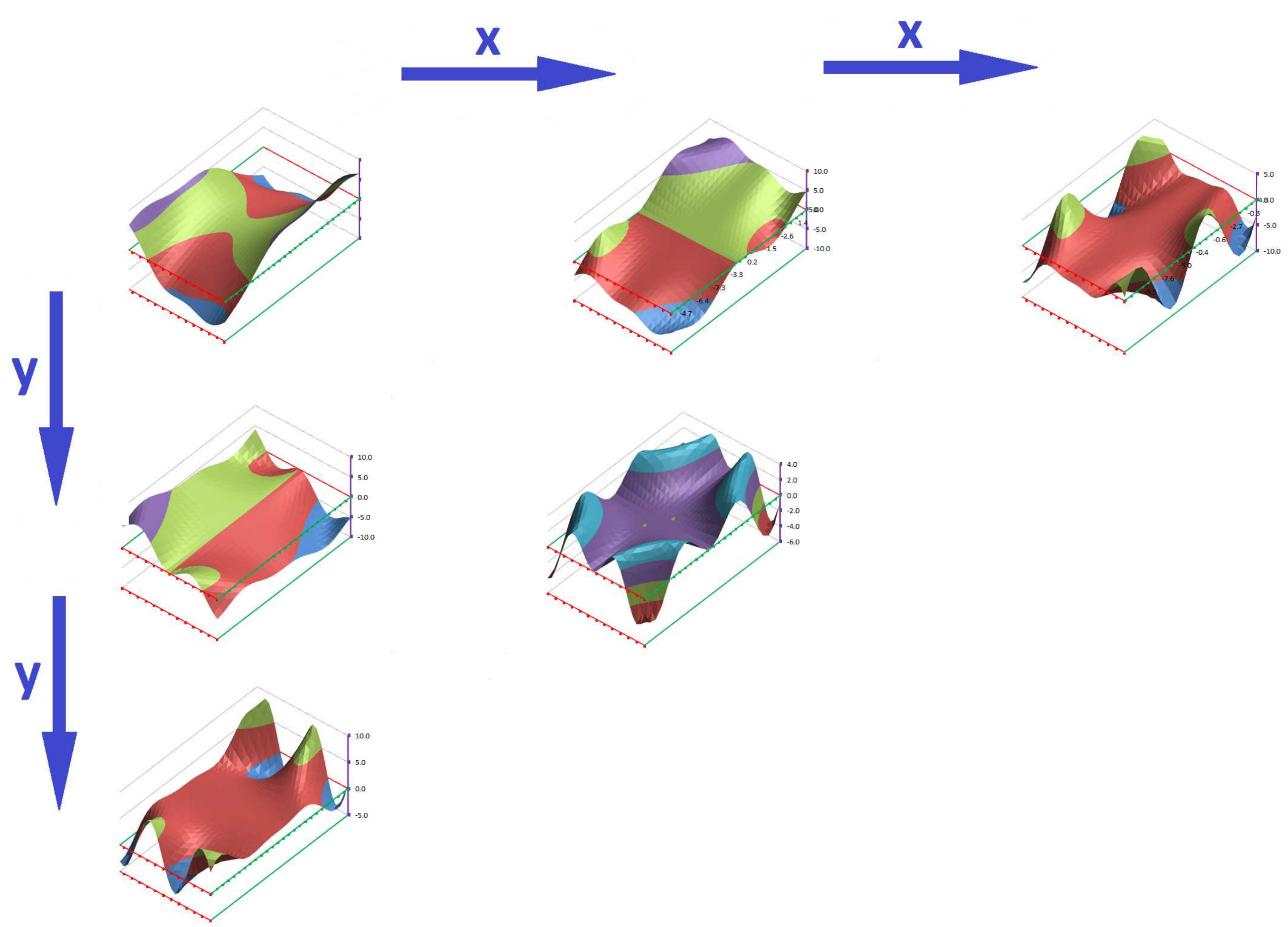


The two functions adjacent to the original are the familiar first partial derivatives. The top right and the bottom left are computed for the first time as described. The function in the middle is discussed in the next section.

The results are all horizontal lines equally spaced and in both cases they form horizontal planes. These planes are the graphs of the two new functions of two variables, the *second partial derivatives*, the tables of which have been constructed. Some things don't change: just as in the one-dimensional case, *the second derivatives of quadratic functions are constant*.

Here is a more complex example:

$$f(x,y) = -x^2 + y^2 + xy + \sin(x+y+3)\,.$$



First, we notice that, just as with numerical functions in Volume 2 (Chapter 2DC-4) and just as with the parametric curves in Chapter 2, the difference of the difference is simply the difference that skips a node. That's why the second difference doesn't provide us – with respect to the same variable – with any meaningful information. We will limit our attention to the *difference quotients.*

We will carry out the same second difference quotient construction for f with respect to x with y fixed and then vice versa. The construction is summarized in this diagram:

$$\begin{array}{ccccccccc}
- & f(x_1) & --- & f(x_2) & --- & f(x_3) & - \\
- & -\bullet- & \dfrac{\Delta f}{\Delta x_2} & -\bullet- & \dfrac{\Delta f}{\Delta x_3} & -\bullet- & - \\
- & -\bullet- & --- & \dfrac{\frac{\Delta f}{\Delta x_3} - \frac{\Delta f}{\Delta x_2}}{s_3 - s_2} & --- & -\bullet- & - \\
 & x_1 & s_2 & x_2 & s_3 & x_3 &
\end{array}$$

The two computations are illustrated by continuing the following familiar diagram for the (first) partial difference quotients:

$$\begin{array}{ccc}
f(x, y+\Delta y) & --- & f(x+\Delta x, y+\Delta y) \\
| & & | \\
| & & | \\
| & & | \\
f(x,y) & --- & f(x+\Delta x, y)
\end{array}
\quad \text{leading to} \quad
\begin{array}{ccc}
-\bullet- & \dfrac{\Delta f}{\Delta x}f(s, y+\Delta y) & -\bullet- \\
| & & | \\
\dfrac{\Delta f}{\Delta y}(x,t) & & \dfrac{\Delta f}{\Delta y}(x+\Delta x, t) \\
| & & | \\
-\bullet- & \dfrac{\Delta f}{\Delta x}(s,y) & -\bullet-
\end{array}$$

There are two further diagrams. We, respectively, subtract the numbers assigned to the horizontal edges horizontally (shown below) and we subtract the numbers assigned to the vertical edges vertically before placing the results at the nodes (not shown):

$$-\bullet- \quad \frac{\Delta f}{\Delta x}(s,y) \quad -\bullet- \quad \frac{\Delta f}{\Delta x}(s+\Delta s, y) \quad -\bullet- \qquad \text{leading to} \qquad -\bullet- \quad -- \quad \frac{\Delta \frac{\Delta f}{\Delta x}}{\Delta x}(x,y) \quad -- \quad -\bullet-$$

Recall from earlier in this chapter that we have a partition P of a rectangle $R = [a,b] \times [c,d]$ in the xy-plane built as a combination of partitions of the intervals $[a,b]$ and $[c,d]$:

$$a = x_0 \le x_1 \le x_2 \le ... \le x_{n-1} \le x_n = b\,,$$

and

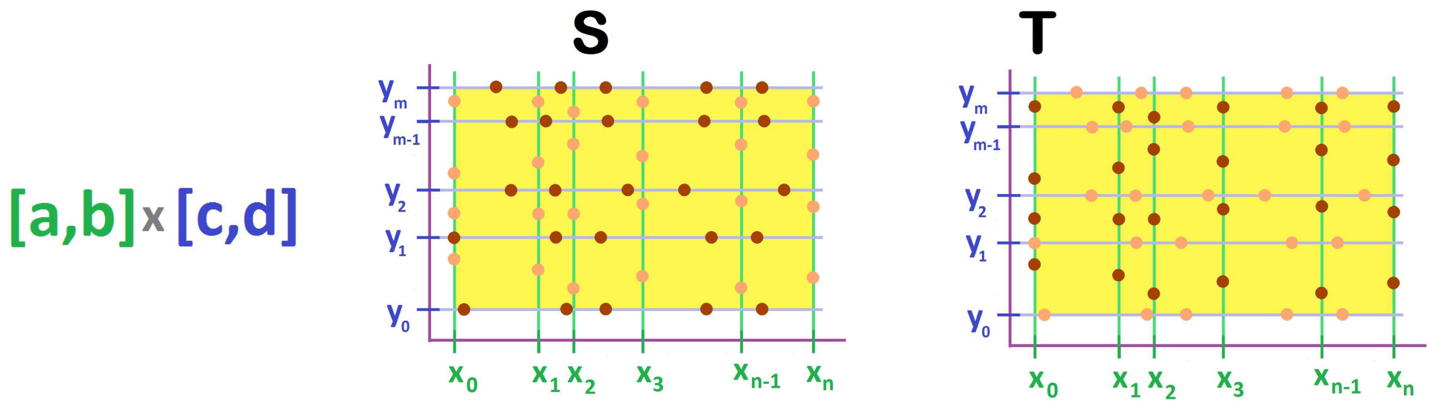
$$c = y_0 \le y_1 \le y_2 \le ... \le y_{m-1} \le y_m = d\,.$$

Its primary nodes are:

$$X_{ij} = (x_i, y_j),\ \ i = 1,2,...,n,\ \ j = 1,2,...,m\,.$$

And its secondary nodes are:

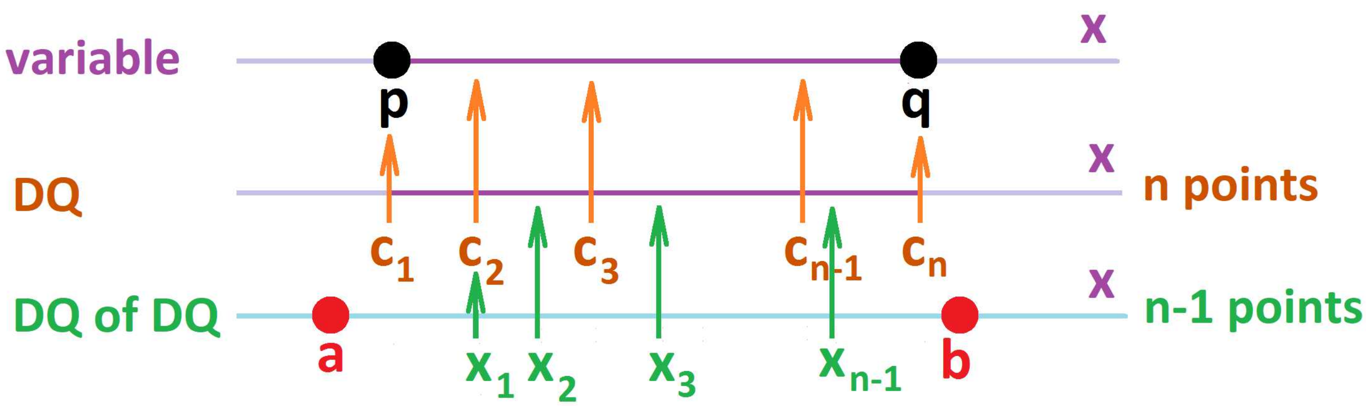
- a point S_{ij} in the segment $[x_{i-1}, x_i] \times \{y_j\}$, and
- a point T_{ij} in the segment $\{x_i\} \times [y_{j-1}, y_j]$.



If $z = f(x, y)$ is defined at the nodes $X_{ij},\ i, j = 0, 1, 2, ..., n$, of the partition, the partial difference quotients of f with respect to x and y are respectively:

$$\frac{\Delta f}{\Delta x}(S_{ij}) = \frac{f(X_{ij}) - f(X_{i-1,j})}{x_i - x_{i-1}} \quad \text{and} \quad \frac{\Delta f}{\Delta y}(T_{ij}) = \frac{f(X_{ij}) - f(X_{i,j-1})}{y_j - y_{j-1}}\,.$$

It is now especially important that we have utilized the secondary nodes as the inputs of the new functions. Indeed, we can now carry out a similar construction with these functions and find their difference quotients, the four of them. The construction is based on the one we carried out for functions of one variable in Volume 2 (Chapter 2DC-3):



We will start with these *two* new quantities.

First, let's consider the rate of change of $\frac{\Delta f}{\Delta x}$ – the rate of change of f with respect to x – with respect to x, again. For each fixed j, we consider the partition of the horizontal interval $[a, b] \times \{y_j\}$ with the primary nodes:

$$S_{ij} = (s_{ij}, y_j),\ \ i = 1, 2, ..., n\,,$$

and the secondary nodes:

$$X_{ij},\ \ i = 1, 2, ..., n - 1\,.$$

We now carry out the same difference quotient construction, still within the horizontal edges.

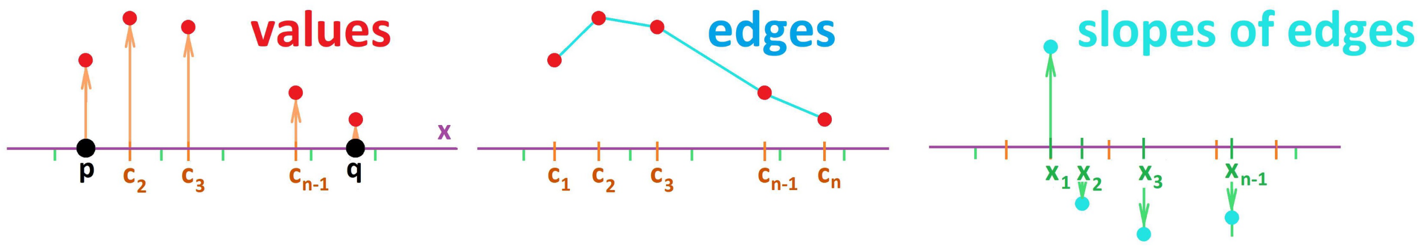
Similarly, to define the rate of change of $\frac{\Delta f}{\Delta y}$ with respect to y. For each fixed i, we consider the partition of the vertical interval $\{x_i\} \times [c, d]$ with the primary nodes:

$$T_{ij} = (x_i, t_{ij}),\ \ j = 1, 2, ..., m\,,$$

and the secondary nodes:

$$X_{ij},\ \ j = 1, 2, ..., m - 1\,.$$

We now carry out the same difference quotient construction, still within the vertical edges.



Definition 3.11.2: second difference quotient

The (repeated) *second difference quotient of f with respect to x* is defined at the primary nodes of the original partition P by the following:

$$\boxed{\frac{\Delta^2 f}{\Delta x^2}(X_{ij}) = \frac{\frac{\Delta f}{\Delta x}(S_{ij}) - \frac{\Delta f}{\Delta x}(S_{i,j-1})}{s_{ij} - s_{i-1,j}}}$$

with $i = 1, 2, ..., n-1,\ j = 0, 1, ..., m$. The *second difference quotient of f with respect to y* is defined at the primary nodes of the original partition P by the following:

$$\boxed{\frac{\Delta^2 f}{\Delta y^2}(X_{ij}) = \frac{\frac{\Delta f}{\Delta y}(T_{ij}) - \frac{\Delta f}{\Delta y}(T_{i,j-1})}{t_{ij} - t_{i,j-1}}}$$

with $i = 0, 1, ..., n,\ j = 1, 2, ..., m-1$.

These are discrete 0-forms.

In the above example, we can see how the sign of these expressions reveal the *concavity* of the curves.

For a simplified notation, we will often *omit the indices*:

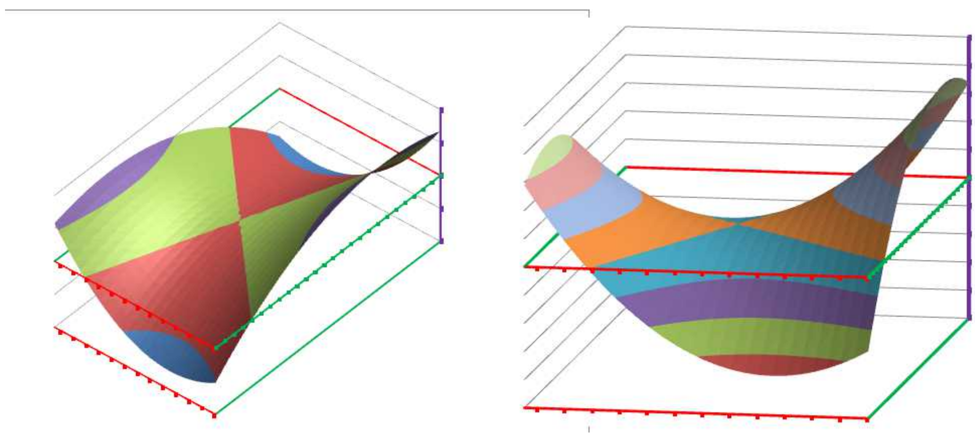
Second difference quotients

$$\frac{\Delta^2 f}{\Delta x^2}(x, y) = \frac{\frac{\Delta f}{\Delta x}(s + \Delta s, y) - \frac{\Delta f}{\Delta x}(s, y)}{\Delta s}$$

$$\frac{\Delta^2 f}{\Delta y^2}(x, y) = \frac{\frac{\Delta f}{\Delta y}(x, t + \Delta t) - \frac{\Delta f}{\Delta y}(x, t)}{\Delta t}$$

3.12. The second difference and the difference quotient with respect to mixed variables

Furthermore, we can compute the rate of change of with respect to the *other* variable. This is how the concavity with respect to x is increasing with increasing y:



We are not in the 1-dimensional setting anymore! That is why, in contrast to the one-dimensional case in Volume 2 (Chapter 2DC-3), the case of parametric curves in Chapter 2, and the repeated variable case above, the difference of the difference is *not* simply the difference that skips a node. The second difference with respect to mixed variables does provide us with meaningful information.

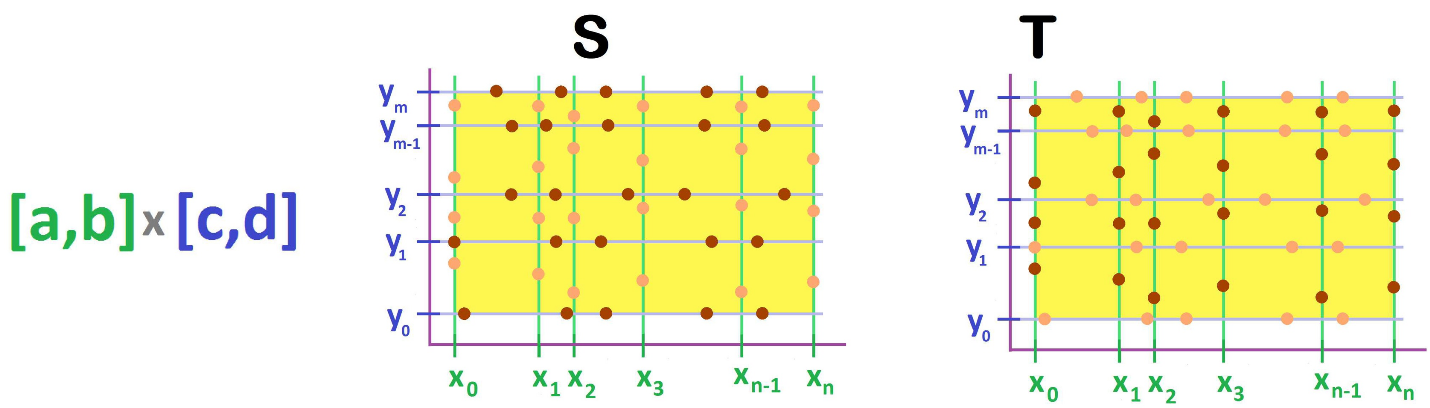
The partial differences:

$$\Delta_x f\,(S_{ij}) \quad \text{and} \quad \Delta_y f\,(T_{ij})\,,$$

and the partial difference quotients:

$$\frac{\Delta f}{\Delta x}(S_{ij}) \quad \text{and} \quad \frac{\Delta f}{\Delta y}(T_{ij})\,.$$

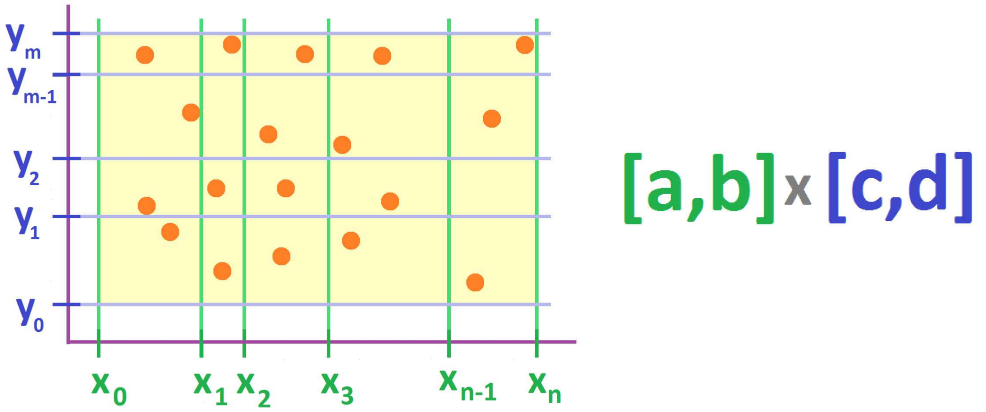
are defined at the secondary nodes located on the edges of the partition of the rectangle $[a,b] \times [c,d]$, horizontal and vertical:



There are four more quantities:

- the change of $\Delta_x f$ with respect to y
- the change of $\Delta_y f$ with respect to x
- the rate of change of $\frac{\Delta f}{\Delta x}$ with respect to y
- the rate of change of $\frac{\Delta f}{\Delta y}$ with respect to x

They will be assigned to the nodes located at the *faces* of the original partition. These are *tertiary nodes*.



The new computations are illustrated by continuing the following familiar diagram for the (first) differences:

$$\begin{array}{ccc}
f(x,y+\Delta y) & --- & f(x+\Delta x,y+\Delta y)\\
| & & |\\
| & & |\\
| & & |\\
f(x,y) & --- & f(x+\Delta x,y)
\end{array}
\quad \text{leading to} \quad
\begin{array}{ccc}
-\bullet- & \Delta_x f\,(s,y+\Delta y) & -\bullet-\\
| & & |\\
\Delta_y f\,(x,t) & & \Delta_y f\,(x+\Delta x,t)\\
| & & |\\
-\bullet- & \Delta_x f\,(s,y) & -\bullet-
\end{array}$$

There are two further diagrams. We, respectively, subtract the numbers assigned to the horizontal edges vertically and we subtract the numbers assigned to the vertical edges horizontally before placing the results in the middle of the square:

$$\text{leading to} \quad
\begin{array}{ccc}
-\bullet- & --- & -\bullet-\\
| & \Delta_y\Delta_x f\,(s,t) & |\\
-\bullet- & --- & -\bullet-
\end{array}
\qquad \text{leading to} \quad
\begin{array}{ccc}
-\bullet- & --- & -\bullet-\\
| & \Delta_x\Delta_y f\,(s,t) & |\\
-\bullet- & --- & -\bullet-
\end{array}$$

Definition 3.12.1: second difference

The (mixed) *second difference of f with respect to yx* is defined at the tertiary nodes of the original partition P by the following:

$$\boxed{\Delta^2_{yx} f\,(U_{ij}) = \Delta_x f\,(S_{ij}) - \Delta_x f\,(S_{i,j-1})}$$

with $i = 1,2,...,n-1,\ \ j = 1,2,...,m-1$. The *mixed difference quotient of f with respect to xy* is defined at the tertiary nodes of the original partition P by the following:

$$\boxed{\Delta^2_{xy} f\,(U_{ij}) = \Delta_y f\,(T_{ij}) - \Delta_y f\,(T_{i-1,j})}$$

with $i = 1,2,...,n-1,\ \ j = 1,2,...,m-1$.

Definition 3.12.2: second difference quotient

The (mixed) *second difference quotient of f with respect to yx* is defined at the tertiary nodes of the original partition P by the following:

$$\boxed{\frac{\Delta^2 f}{\Delta y \Delta x}(U_{ij}) = \frac{\frac{\Delta f}{\Delta x}(S_{ij}) - \frac{\Delta f}{\Delta x}(S_{i,j-1})}{y_j - y_{j-1}}}$$

with $i = 1,2,...,n-1,\ \ j = 1,2,...,m-1$. The *mixed difference quotient of f*

with respect to xy is defined at the tertiary nodes of the original partition P by the following:

$$\boxed{\frac{\Delta^2 f}{\Delta x \Delta y}(U_{ij}) = \frac{\frac{\Delta f}{\Delta y}(T_{i,j}) - \frac{\Delta f}{\Delta y}(T_{i-1,j})}{x_i - x_{i-1}}}$$

with $i = 1, 2, ..., n-1,\ j = 1, 2, ..., m-1$.

When the tertiary nodes are unspecified, these are functions of the faces themselves, i.e., discrete 2-forms.

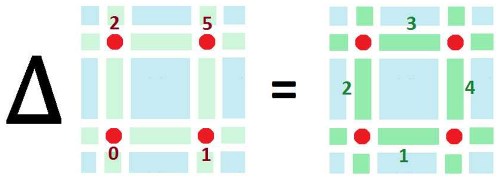
For a simplified notation, we will often *omit the indices*:

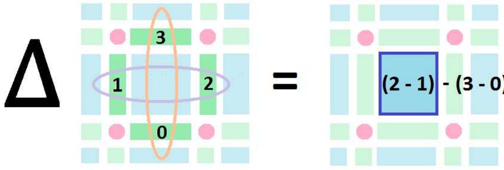
Second difference quotients

$$\frac{\Delta^2 f}{\Delta y \Delta x}(s,t) = \frac{\frac{\Delta f}{\Delta x}(s, y+\Delta y) - \frac{\Delta f}{\Delta x}(s,y)}{\Delta y}$$

$$\frac{\Delta^2 f}{\Delta x \Delta y}(s,t) = \frac{\frac{\Delta f}{\Delta y}(x+\Delta x, t) - \frac{\Delta f}{\Delta y}(x,t)}{\Delta x}$$

Now, the two mixed differences are "made of" the same four quantities; one can guess that they are equal. For example, we can see it here:





Let's confirm that with algebra. We substitute and simplify:

$$\begin{aligned}
\Delta^2_{yx} f\,(s,t) &= \Delta_x f\,(s, y+\Delta y) - \Delta_x f\,(s,y) \\
&= \big(f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y)\big) - \big(f(x+\Delta x, y) - f(x,y)\big) \\
&= f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y) - f(x+\Delta x, y) + f(x,y) \\
&= \big(f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y)\big) - \big(f(x, y+\Delta y) - f(x,y)\big) \\
&= \Delta_y f\,(x+\Delta x, t) - \Delta_y f\,(x,t) \\
&= \Delta^2_{xy} f\,(s,t)\,.
\end{aligned}$$

Theorem 3.12.3: Discrete Clairaut's Theorem

Over a partition in $\mathbf{R}^n$, first, the mixed second differences with respect to any two variables are equal to each other:

$$\boxed{\Delta^2_{yx} f = \Delta^2_{xy} f}$$

and, second, the mixed second difference quotients are equal to each other:

$$\boxed{\frac{\Delta^2 f}{\Delta y \Delta x} = \frac{\Delta^2 f}{\Delta x \Delta y}}$$

This theorem will have important consequences presented in Chapter 4.

3.13. The second partial derivatives

The *second derivative* is known to help to classify extreme points. At least, we can dismiss the possibility that a point is a maximum when the function is concave up, i.e., the second derivative is positive.

Example 3.13.1: a computation

In the above example of the function

$$f(x,y) = x^2 + 3y^2 + xy + 7\,,$$

we differentiate one more time each partial derivative – with respect to the same variable:

$$\begin{array}{llll}
\dfrac{\partial f}{\partial x} & = 2x + y & \Longrightarrow & \dfrac{\partial}{\partial x}\left(\dfrac{\partial f}{\partial x}\right) & = \dfrac{\partial}{\partial x}(2x+y) & = 2 \\
\dfrac{\partial f}{\partial x} & = 6y + x & \Longrightarrow & \dfrac{\partial}{\partial y}\left(\dfrac{\partial f}{\partial y}\right) & = \dfrac{\partial}{\partial x}(6y+x) & = 6
\end{array}$$

Both numbers are positive, therefore, both curves are concave up!

Repeated differentiation produces a sequence of functions in the one-variable case:

$$f \xrightarrow{\frac{d}{dx}} f' \xrightarrow{\frac{d}{dx}} f'' \xrightarrow{\frac{d}{dx}} \ldots \xrightarrow{\frac{d}{dx}} f^{(n)} \xrightarrow{\frac{d}{dx}} \ldots$$

In the two-variable case, it makes the functions *multiply*:

$$\begin{tikzcd}[column sep=small]
 & & & f \arrow[dll, "x"'] \arrow[drr, "y"] & & & \\
 & f_x \arrow[dl, "x"'] \arrow[dr, "y"] & & & & f_y \arrow[dl, "x"'] \arrow[dr, "y"] & \\
f_{xx} \arrow[d, "x"'] \arrow[dr, "y"] & & f_{yx} \arrow[d, "x"'] & & f_{xy} \arrow[d, "y"] & & f_{yy} \arrow[d, "x"'] \arrow[dr, "y"] & \\
\ldots & \ldots & \ldots & & \ldots & & \ldots & \ldots
\end{tikzcd}$$

The number of derivatives might be reduced when the *mixed derivatives* at the bottom are equal. It is often possible thanks to the following result that we accept without proof.

Theorem 3.13.2: Clairaut's Theorem

Suppose a function $z = f(x,y)$ has continuous second partial derivatives at a

given point (x_0, y_0) in $\mathbf{R}^2$. Then we have:

$$\boxed{\frac{\partial^2 f}{\partial x\,\partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y\,\partial x}(x_0, y_0)}$$

Under the conditions of the theorem, this part of the above diagram becomes commutative:

$$\begin{array}{ccccccc} & & & f & & & \\ & & \swarrow_x & & {}_y\searrow & & \\ & f_x & & & & f_y & \\ \swarrow_x & & {}_y\searrow & & \swarrow_x & & {}_y\searrow \\ f_{xx} & & & f_{yx}=f_{xy} & & & f_{yy} \end{array}$$

In fact, *the two operations of partial differentiation commute*:

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x}\,.$$

This can also be written as another commutative diagram:

$$\begin{tikzcd} \boxed{f} \arrow[r, "\frac{\partial}{\partial x}"] \arrow[d, "\frac{\partial}{\partial y}"] \arrow[dr] & f_x \arrow[d, "\frac{\partial}{\partial y}"] \\ f_y \arrow[r, "\frac{\partial}{\partial x}"] & f_{xy}=f_{yx} \end{tikzcd}$$

For the case of *three* variables, there are even more derivatives:

$$\begin{array}{ccccccccc} & & & & \boxed{f} & & & & \\ & \swarrow_x & & & \downarrow_y & & & {}_z\searrow & \\ & f_x & & & f_y & & & f_z & \\ \swarrow_x & \downarrow_y & {}_z\searrow & \swarrow_x & \downarrow_y & {}_z\searrow & \swarrow_x & \downarrow_y & {}_z\searrow \\ f_{xx} & f_{yx} & f_{zx} & f_{xy} & f_{yy} & f_{zy} & f_{xz} & f_{xz} & f_{yz}\ f_{zz} \end{array}$$

Under the conditions of the theorem, *the three operations of partial differentiation commute* (pairwise):

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x}, \quad \frac{\partial}{\partial y}\frac{\partial}{\partial z} = \frac{\partial}{\partial z}\frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z}\frac{\partial}{\partial x} = \frac{\partial}{\partial x}\frac{\partial}{\partial z}$$

The above diagram also becomes commutative:

$$\begin{array}{ccccccc} & & & f_{zz} & & & \\ & & & \uparrow_z & & & \\ & \longleftarrow & \longleftarrow & f_z & \longrightarrow & \longrightarrow & \\ \swarrow^x & & & \uparrow_z & & & {}^y\searrow \\ f_{xz}= & f_{zx} & & \boxed{f} & & f_{zy} & =f_{yz} \\ & \uparrow_z & \swarrow^x & & {}^y\searrow & \uparrow_z & \\ & f_x & & & & f_y & \\ \swarrow^x & & {}^y\searrow & & \swarrow^x & & {}^y\searrow \\ f_{xx} & & & f_{yx}=f_{xy} & & & f_{yy} \end{array}$$

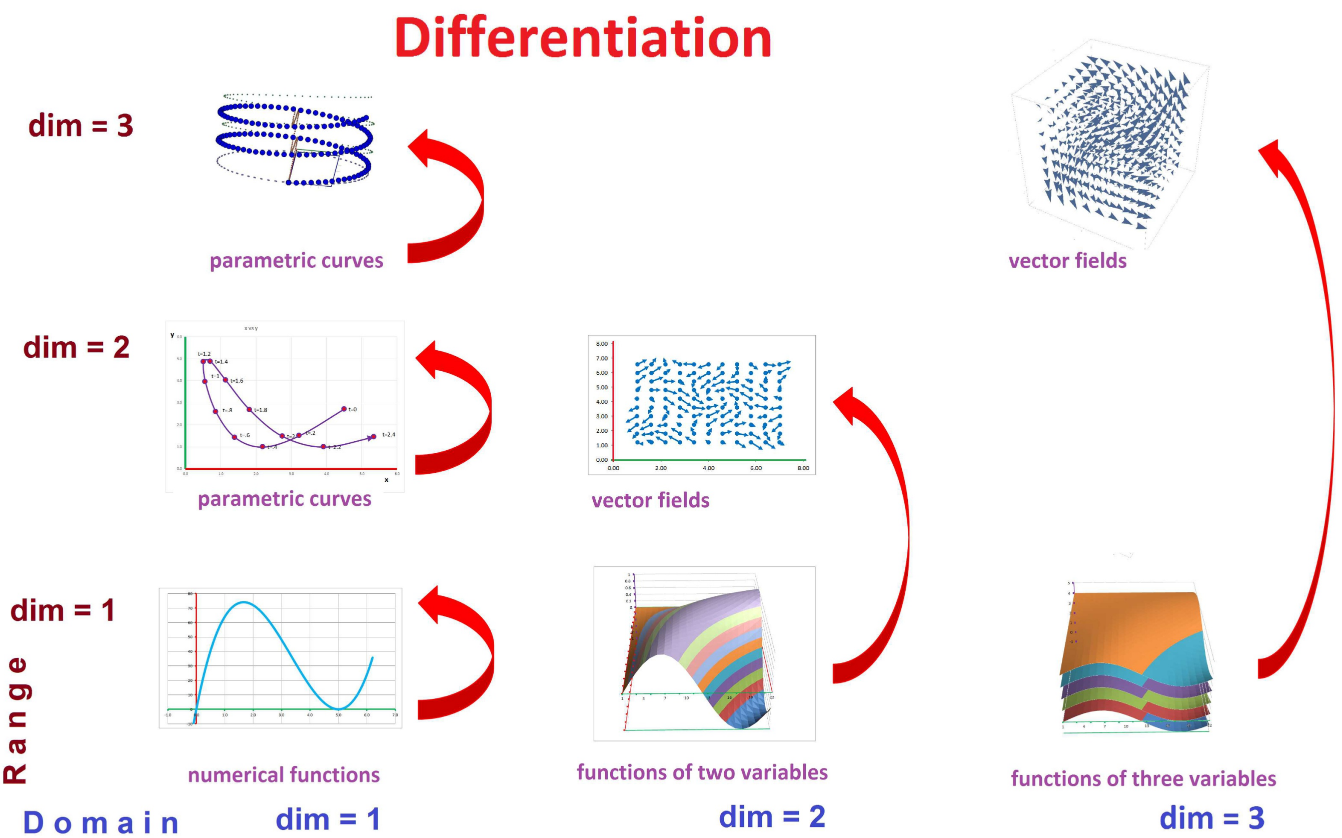
Chapter 4: The gradient

Contents

4.1. Overview of differentiation

Where are we in our study of functions in high dimensions?

Once again, we provide a diagram that captures all types of functions we have seen so far as well as those haven't seen yet. They are placed on the xy-plane with the x-axis and the y-axis representing the dimensions of the input space and the output space. The first column consists of all parametric curves and the first row of all functions of several variables. The two have one cell in common; that is numerical functions.

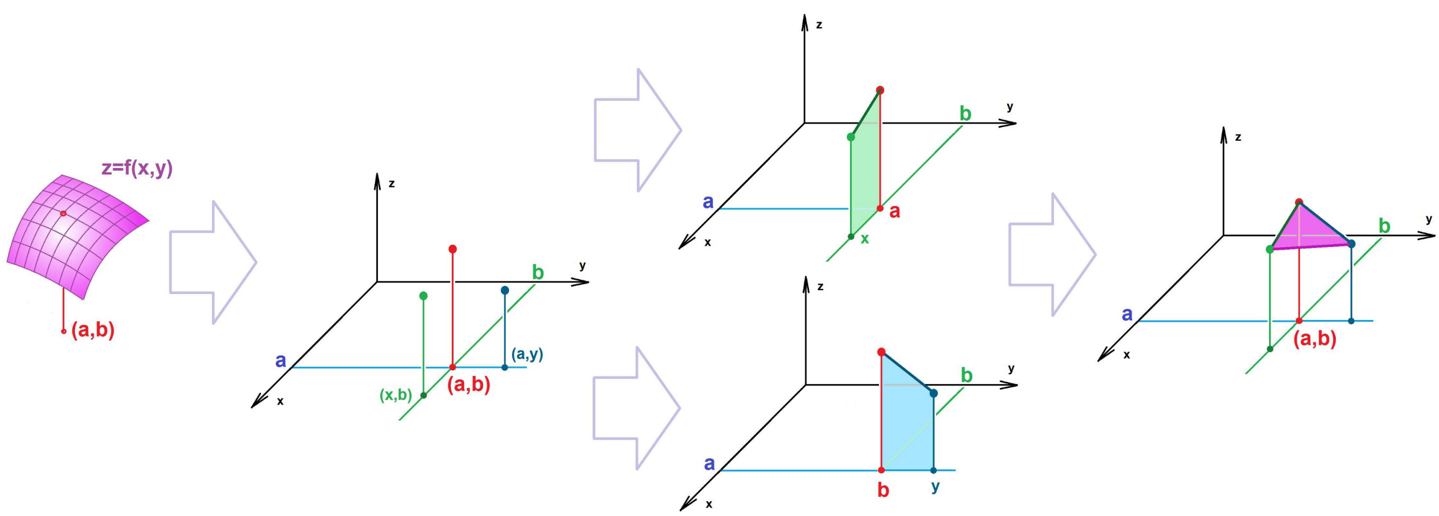


This time we will see how everything is interconnected. We show with the red arrows for different types of functions what type of functions are their difference quotients or the derivatives.

In Volume 2 (Chapter 2DC-3) and beyond, we faced only numerical functions and we implicitly used the fact that differentiation will not make us leave the confines of this environment. In particular, every function defined on the edges of a partition is the difference quotient of some function defined on the nodes of the partition. Furthermore, every continuous function is integrable and, therefore, is somebody's derivative. In this sense, the arrow can be reversed.

More recently, we defined the difference quotient and the derivative of a parametric curve in $\mathbf{R}^n$ and those are again parametric curves in $\mathbf{R}^n$ (think location vs. velocity). That's why we have arrows that come back to the same cell. Once again, every continuous parametric curve is somebody's derivative. In this sense, the arrow can be reversed.

The study of functions of several variables would be incomplete without understanding their rates of change! What we know so far is that we can compute the rate of change of such a function of two variables in the two main directions:



The result is given by a vector called its *gradient*.

For each point on the plane, we have a single vector but what if we carry out this computation over the whole plane? What if, in order to keep track of the correspondence, we attach this vector to the point it came from? The result is a *vector field*. It is a function from $\mathbf{R}^2$ to $\mathbf{R}^2$ and it is placed on the diagonal of our table. The same happens to functions of three variables and so on.

Every gradient is a vector field but not every vector field is a gradient. In this sense, the arrow cannot we reversed! The situation seems to mimic the one with numerical functions: there are non-integrable functions. However, the arrow in the first cell *is* reversible if we limit ourselves to smooth (i.e., infinitely many times differentiable) functions. The problem is more profound with vector fields as we shall see later.

4.2. Gradients vs. vector fields

Vector fields are just functions with:

- the input consisting of two numbers and
- the output consisting of two numbers.

We just choose to treat the former as a point on the plane and the latter as a vector attached to that point. This is just a clever way to visualize such a complex – in comparison to the ones we have seen so far – function. It's a *location-dependent vector*!

Let's plot some vector fields *by hand* and then analyze them.

Example 4.2.1: accelerated outflow

Let's consider this simple vector field:

$$V(x,y) =< x, y > \ .$$

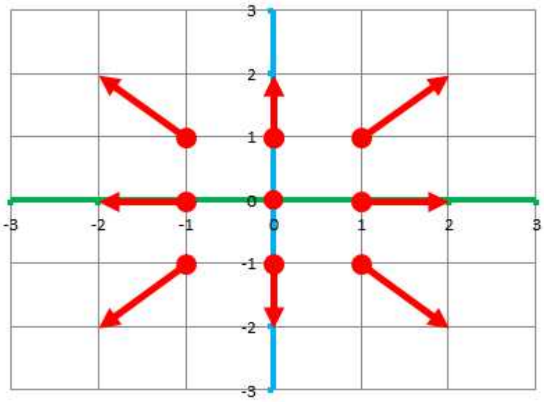
A vector field is just two functions of two variables:

$$V(x,y) =< p(x,y),\ q(x,y) > \text{ with } p(x,y) = x,\ q(x,y) = y\,.$$

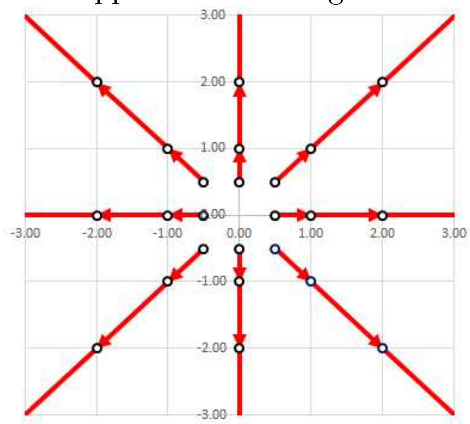
Plotting those two functions, as before, does not produce useful visualization. However, we will still follow the same pattern: we pick a few points on the plane, compute the output for each, and assign it to that point. The difference is that instead of a single number we have two and instead of a vertical bar that we erect at that point to visualize this number we draw an arrow. We carry this out for these nine points around the origin:

$$\begin{array}{ccc} (-1,1) & (0,1) & (1,1) \\ (-1,0) & (0,0) & (1,0) \\ (-1,-1) & (0,-1) & (1,-1) \end{array} \quad \text{leading to} \quad \begin{array}{ccc} < -1,1 > & < 0,1 > & < 1,1 > \\ < -1,0 > & < 0,0 > & < 1,0 > \\ < -1,-1 > & < 0,-1 > & < 1,-1 > \end{array}$$

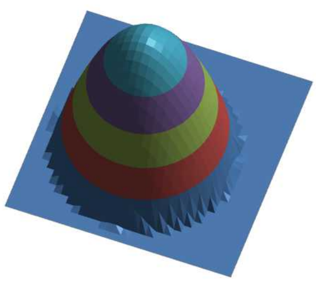
Each vector on right starts at the corresponding point on left:

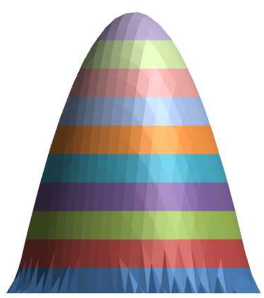


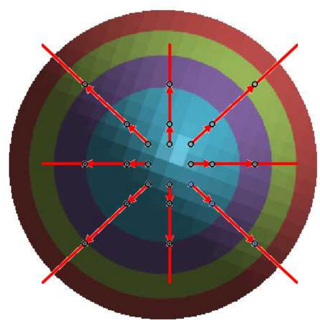
What about the rest? We can guess that the magnitudes increase as we move away from the origin while the directions remain the same: opposite of the origin.



For each point (x, y) we copy the vector that ends there, i.e., $< x, y >$ and place it at this location. Now, if the vectors represent velocities of particles, what kind of flow is this? This isn't a fountain or an explosion (the particles would go slower away from the source). The answer is: this is a *flow on a surface* – under gravity – that gets steeper and steeper away from the origin! This surface would look like this paraboloid:







Can we be more specific? Well, this surface looks like the graph of a function of two variables and the flow seems to follow the line fastest descent; maybe our vector field is the gradient of this function? We will find out but first let's take a look at the vector field visualized as a system of pipes:

	x		-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
y																							
4.0			-1.5		-0.5		0.5		1.5		2.5		3.5		4.5		5.5		6.5		7.5		8.5
3.5		4		4		4		4		4		4		4		4		4		4		4	
3.0			-1.5		-0.5		0.5		1.5		2.5		3.5		4.5		5.5		6.5		7.5		8.5
2.5		3		3		3		3		3		3		3		3		3		3		3	
2.0			-1.5		-0.5		0.5		1.5		2.5		3.5		4.5		5.5		6.5		7.5		8.5
1.5		2		2		2		2		2		2		2		2		2		2		2	
1.0			-1.5		-0.5		0.5		1.5		2.5		3.5		4.5		5.5		6.5		7.5		8.5
0.5		1		1		1		1		1		1		1		1		1		1		1	
0.0			-1.5		-0.5		0.5		1.5		2.5		3.5		4.5		5.5		6.5		7.5		8.5
-0.5		-1		-1		-1		-1		-1		-1		-1		-1		-1		-1		-1	

We recognize this as a discrete 1-form. Now, the question above becomes: is it possible to produce this pattern of flow in the pipes by controlling the pressure at the joints? So, is this vector field – when limited to the edges of the grid – the difference of a function of two variables?

Let's take the latter question; we need to solve this equation and such a function $z = f(x, y)$ that

$$\Delta f =< x, y > .$$

The latter is just an *abbreviation*; the actual differences are x on the horizontal edges and y on the vertical. Let's concentrate on just one cell:

$$[x, x + \Delta x] \times [y, y + \Delta y] .$$

We choose the secondary node of an edge to be the primary node at the beginning of the edge:

- horizontal: (x, y) in $[x, x + \Delta x] \times \{y\}$, etc.
- vertical: (x, y) in $\{x\} \times [y, y + \Delta y]$, etc.

Then our equation develops as follows:

$$\Longrightarrow \begin{cases} \Delta_x f &= x \\ \Delta_x f &= y \end{cases} \Longrightarrow \begin{cases} f(x + \Delta x, y) - f(x, y) &= x \\ f(x, y + \Delta y) - f(x, y) &= y \end{cases} \Longrightarrow \begin{cases} f(x + \Delta x, y) &= f(x, y) + x \\ f(x, y + \Delta y) &= f(x, y) + y \end{cases}$$

These recursive relations allow us to construct f one node at a time. We use the first one to progress horizontally and the second to progress vertically:

$$\begin{array}{ccccc} \uparrow & & \uparrow & & \\ (x, y + \Delta y) & \to & (x + \Delta x, y + \Delta y) & \to \\ \uparrow & & \uparrow & & \\ (x, y) & \to & (x + \Delta x, y) & \to \end{array}$$

The problem is solved! However, there may be a *conflict*: what if we apply these two formulas consecutively but in a different order? Fortunately, going horizontally then vertically produces the same outcome as going vertically then horizontally:

$$f(x + \Delta x, y + \Delta y) = f(x, y) + x + y\,.$$

This is the first instance of *path-independence*.

Now the continuous case. Suppose V is the gradient of some differentiable function of two variables $z = f(x, y)$. The result of this assumption is a vector equation that breaks into two:

$$V = \nabla f \Longrightarrow V(x, y) =< x, y >=< f_x(x, y)\,, f_y(x, y) > \Longrightarrow \begin{cases} x &= f_x(x, y), \\ y &= f_y(x, y). \end{cases}$$

We now integrate one variable at a time:

$$\begin{array}{llllll} x &= f_x(x, y) & \Longrightarrow & f(x, y) = \displaystyle\int x\,dx & = \dfrac{x^2}{2} + C & = \dfrac{x^2}{2} + C(y) \\ y &= f_y(x, y) & \Longrightarrow & f(x, y) = \displaystyle\int y\,dy & = \dfrac{y^2}{2} + K & = \dfrac{y^2}{2} + K(x) \end{array}$$

Note that in either case we add the familiar constants of integration "$+C$" and "$+K$" (different for the two different integrations)... however, these constants are only constant relative to x and y respectively. That makes them *functions* of y and x respectively! Putting the two together, we have the following restriction on the two unknown functions:

$$f(x, y) = \frac{x^2}{2} + C(y) = \frac{y^2}{2} + K(x)\,.$$

Can we find such functions C and K? If we *group* the terms, the choice becomes obvious:

$$\begin{array}{ccc} \dfrac{x^2}{2} & & K(x) \\ & = & \\ +C(y) & & +\dfrac{y^2}{2} \end{array}$$

If (or when) it does not, we could just plug some values into this equation and examine the results:

$$\begin{array}{lllll} x = 0 & \Longrightarrow & C(y) = \dfrac{y^2}{2} + K(0) & \Longrightarrow & C(y) = \dfrac{y^2}{2} + \text{ constant} \\ y = 0 & \Longrightarrow & C(0) + \dfrac{x^2}{2} = K(x) & \Longrightarrow & K(x) = \dfrac{x^2}{2} + \text{ constant} \end{array}$$

So, either of the functions $C(y)$ and $K(x)$ differs from the corresponding expression by a constant. Therefore, we have:

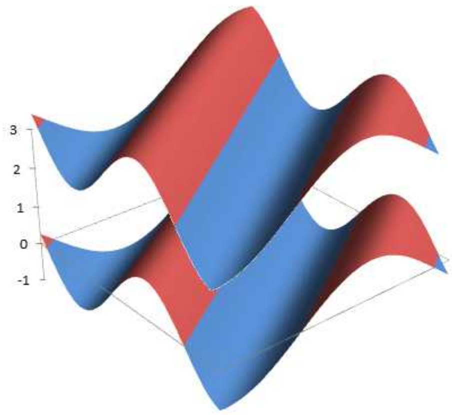
$$f(x,y)=\frac{x^2}{2}+\frac{y^2}{2}+L\,,$$

for some constant L. The surface is indeed a paraboloid of revolution.

When a vector field is the gradient of some function of two variables, this function is called a *potential function*, or simply a potential, of the vector field.

Note that finding for a given vector field V a function f such that $\nabla f=V$ amounts to *anti-differentiation*.

The following is an analog of several familiar results: any two potential functions of the same vector field defined on an open disk differ by a constant. So, *you've found one – you've found all*, just like in Volume 2. The proof is exactly the same as before but it relies, just as before, on the properties of the derivatives (i.e., gradients) discussed in the next section. The graphs of these functions then differ by a vertical shift.



It is as if the floor and the ceiling in a cave have the exact same slope in all directions at each location; then the height of the ceiling is the same throughout the cave.

Example 4.2.2: rotational flow

We consider this vector field again:

$$V(x,y)=<y\,,-x>\ .$$

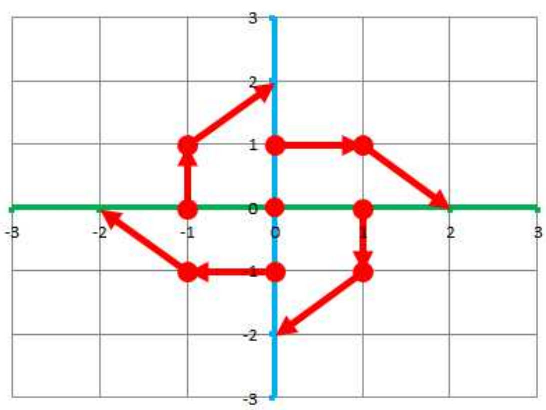
Our two functions of two variables are:

$$V(x,y)=<p(x,y)\,,q(x,y)>\ \text{ with } p(x,y)=y,\ q(x,y)=-x\,.$$

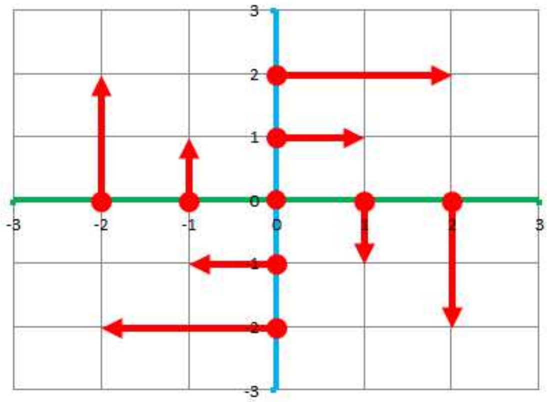
We pick a few points on the plane, compute the output for each, and assign it to that point:

$$\begin{array}{ccccccc}(-1,1)&(0,1)&(1,1)&&<1,-1>&<1,0>&<1,-1>\\(-1,0)&(0,0)&(1,0)&\text{leading to}&<0,1>&<0,0>&<0,-1>\\(-1,-1)&(0,-1)&(1,-1)&&<-1,1>&<-1,0>&<-1,1>\end{array}$$

Each vector on right starts at the corresponding point on left:

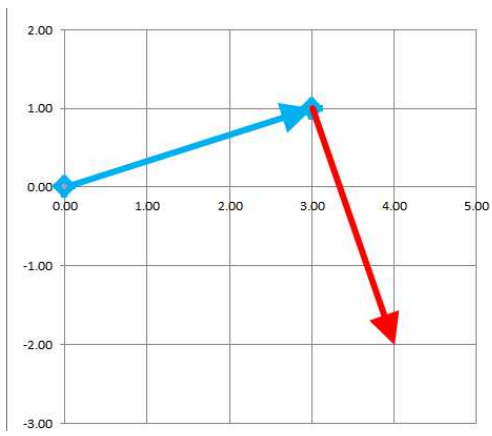


Now, if the vectors represent velocities of particles, what kind of flow is this? It looks like the water is flowing away from the center. Is it a *whirl*? Let's plot some more:



These lie on the axes and they are all *perpendicular* to those axes. We realize that there is a pattern: $V(x,y)$ is perpendicular to $< x, y >$. Indeed,

$$< y, -x > \cdot < x, y >= yx - xy = 0\,.$$



From what we know about parametric curves, to follow these arrows a curve would be rounding the origin never getting closer to or farther away from it; this must be a *rotation*. Now, is this a *flow on a surface* produced by gravity like last time? If we visualize the vector field as a system of pipes, the question above becomes: is it possible to produce this pattern of flow in the pipes by controlling the pressure at the joints?

	x		-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
y																							
4.0			4		4		4		4		4		4		4		4		4		4		4
3.5		2		1		0		-1		-2		-3		-4		-5		-6		-7		-8	
3.0			3		3		3		3		3		3		3		3		3		3		3
2.5		2		1		0		-1		-2		-3		-4		-5		-6		-7		-8	
2.0			2		2		2		2		2		2		2		2		2		2		2
1.5		2		1		0		-1		-2		-3		-4		-5		-6		-7		-8	
1.0			1		1		1		1		1		1		1		1		1		1		1
0.5		2		1		0		-1		-2		-3		-4		-5		-6		-7		-8	
0.0			0		0		0		0		0		0		0		0		0		0		0
-0.5		2		1		0		-1		-2		-3		-4		-5		-6		-7		-8	

Let's find out. We will try to solve this equation for $z = f(x,y)$:

$$\Delta f =< y\,, -x > \ .$$

Just as in the last example, we choose the secondary node to be the primary node at the beginning of the edge. We have:

$$\Longrightarrow \begin{cases} \Delta_x f &= y \\ \Delta_y f &= -x \end{cases} \quad \Longrightarrow \begin{cases} f(x+\Delta x, y) - f(x,y) &= y \\ f(x, y+\Delta y) - f(x,y) &= -x \end{cases} \quad \Longrightarrow \begin{cases} f(x+\Delta x, y) &= f(x,y) + y \\ f(x, y+\Delta y) &= f(x,y) - x \end{cases}$$

Can we use these recursive formulas to construct f? Is there a *conflict*: if we start at (x,y) and then get to $(x+\Delta x, y+\Delta y)$ in the two different ways, will we have the same outcome?

$$\begin{array}{ccccc}
\uparrow & & \uparrow & & \\
(x,y+\Delta y) & \to & (x+\Delta x,y+\Delta y) & \to & \\
\uparrow & & \uparrow & & \\
(x,y) & \to & (x+\Delta x,y) & \to &
\end{array}$$

Unfortunately, the outcome is *not* the same:

$$\begin{array}{rlll}
 & f(x+\Delta x,y+\Delta y) & =f(x,y)+y-(x+\Delta x) & =f(x,y)-x+y-\Delta x\\
\neq & f(x+\Delta x,y+\Delta y) & =f(x,y)-x+(y+\Delta y) & =f(x,y)-x+y+\Delta y
\end{array}$$

This is *path-dependence*!

Suppose V is the gradient of some function of two variables f:

$$V=\nabla f \implies V(x,y)=< y\,,-x >=< f_x(x,y)\,,f_y(x,y) > \implies \begin{cases} y & =f_x(x,y),\\ -x & =f_y(x,y).\end{cases}$$

What do we do with those? They are partial derivatives so let's solve these equations by partial integration, one variable at a time:

$$\begin{array}{llll}
y & =f_x(x,y) & \implies f(x,y)=\displaystyle\int y\,dx & =xy+C\\
-x & =f_y(x,y) & \implies f(x,y)=\displaystyle\int -x\,dy & =-xy+K
\end{array}$$

Putting the two together, we have:

$$f(x,y)=xy+C(y)=-xy+K(x)\,.$$

Can we find such functions C and K? If we try to *group* the terms, they don't group well:

$$\begin{array}{ccc}
xy & & -xy\\
 & = & \\
+C(y) & & +K(x)
\end{array}$$

To confirm that there is a problem, let's plug some values into this equation:

$$\begin{array}{lllll}
x=0 & \implies & C(y)=K(0) & \implies & C(y)=\text{ constant}\\
y=0 & \implies & C(0)=K(x) & \implies & K(x)=\text{ constant}
\end{array}$$

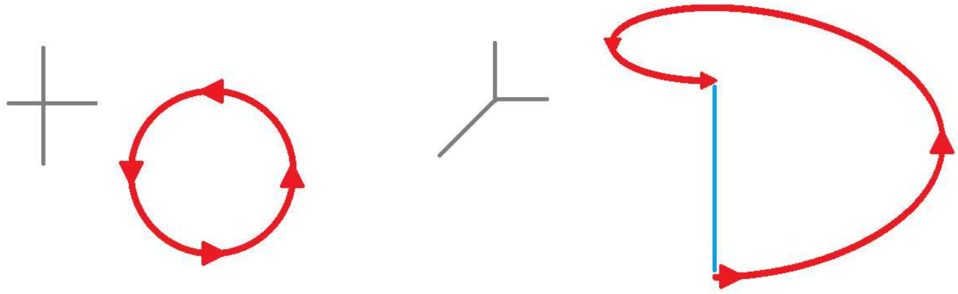
So, both C and K are constant functions, which is impossible! Indeed, on left we have a function of two variables and a constant on right:

$$2xy=-C+K=\text{ constant}\,.$$

This contradiction proves that our assumption that V has a potential function was wrong; there is no such f. We may even say that the vector field isn't "integrable"! Geometrically, there is no surface a flow of water on which would produce this pattern.

Example 4.2.3: climbing

An insightful if informal argument to the same effect is as follows. Suppose we travel along the arrows of a vector field. Suppose that eventually we arrive to our original location. Is it possible that this vector field has a potential function? Is it the gradient of some function of two variables? If it is, we have followed the direction of the (fastest) increase of this function... but once we have come back, what is the elevation? After all this climbing, it can't be the same:



This function therefore cannot be continuous at this location. Then it also cannot be differentiable, a contradiction!

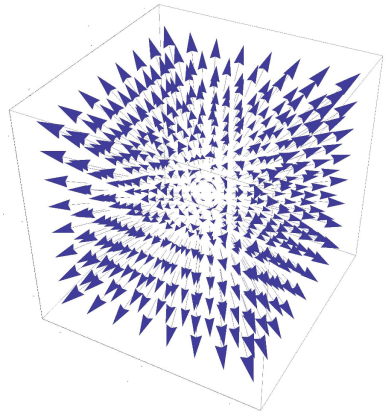
Our conclusion that some continuous vector fields on the plane aren't derivatives has no analog in the 1-dimensional, numerical, case discussed in Volumes 2 and 3.

Example 4.2.4: source

Three-dimensional vector fields are more complex. The one below is similar to the first example above:

$$V(x,y,z) = < x, y, z > \ .$$

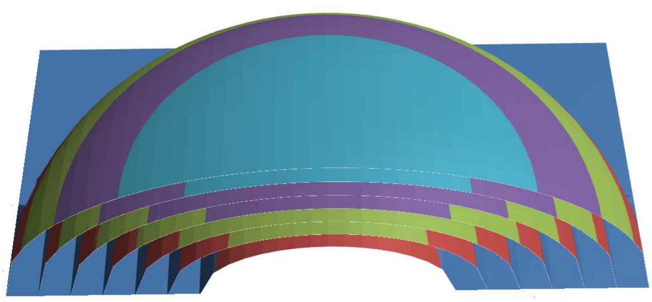
The vectors point in the direction opposite of the direction to the origin:



Just as its two-dimensional analog, there is a potential function:

$$f(x,y,z) = \frac{x^2}{2} + \frac{y^2}{2} + \frac{z^2}{2} \, .$$

The level surfaces of this function are concentric spheres:



4.3. The change and the rate of change of a function of several variables

We consider functions of n variables again.

First let's look at the point-slope form of *linear functions*:

$$l(x_1, ..., x_n) = p + m_1(x_1 - a_1) + ... + m_n(x_n - a_n)\,,$$

where p is the z-intercept, $m_1, ..., m_n$ are the chosen slopes of the plane along the axes, and $a_1, ..., a_n$ are the coordinates of the chosen point in $\mathbf{R}^n$. Let's recast this expression, as before, in terms of the *dot product* with the increment of the independent variable:

$$\boxed{l(X) = p + M \cdot (X - A)}$$

Here we have:

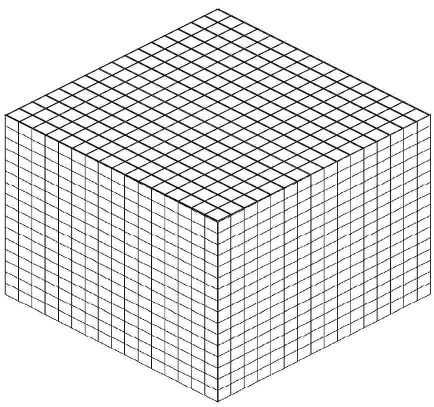
- $M =< m_1, ..., m_n >$ is the vector of slopes.
- $A = (a_1, ..., a_n)$ is the point in $\mathbf{R}^n$.
- $X = (x_1, ..., x_n)$ is our variable point in $\mathbf{R}^n$.
- $X - A$ is how far we step away from our point of interest A.

Then we can say that the vector $N =< m_1, ..., m_n, 1 >$ is perpendicular to the graph of this function (a "plane" in $\mathbf{R}^{n+1}$). The conclusion holds independently from any choice of a coordinate system!

Suppose now we have an arbitrary function $z = f(X)$ of n variables, i.e., z is a real number and $X = (x_1, ..., x_n)$ in $\mathbf{R}^n$.

We start with the discrete case.

Suppose $\mathbf{R}^n$ is equipped with a rectangular grid with sides Δx_k along the axis of each variable x_k. It serves as a partition with secondary nodes provided on the edges of the grid.



The definition of the difference is exactly the same as before:

Definition 4.3.1: difference

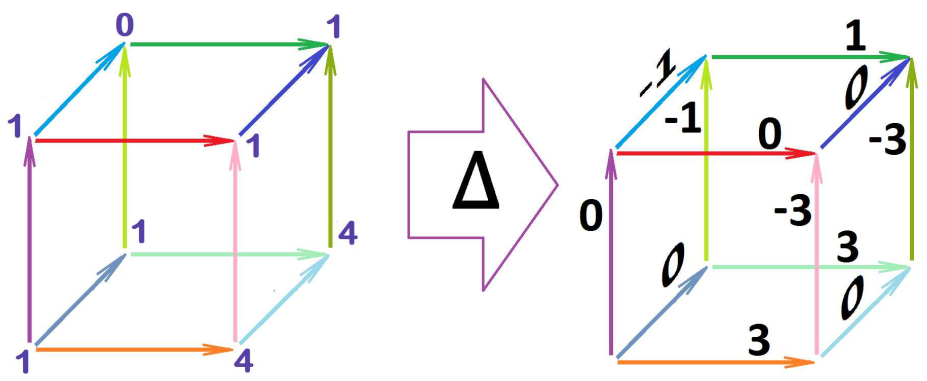
The *difference* of $z = f(X) = f(x_1, ..., x_n)$ at C is defined to be the change of z

with respect to the increment ΔX denoted as follows:

$$\boxed{\Delta f\,(C) = f(X + \Delta X) - f(X)}$$

where C is the secondary node of the edge between X and $X + \Delta X$.

What used to be the difference of the values of the function at two consecutive noes has become the difference of the values of the function at two *adjacent* nodes:



Each difference is computed from tip to toe of the oriented edge.

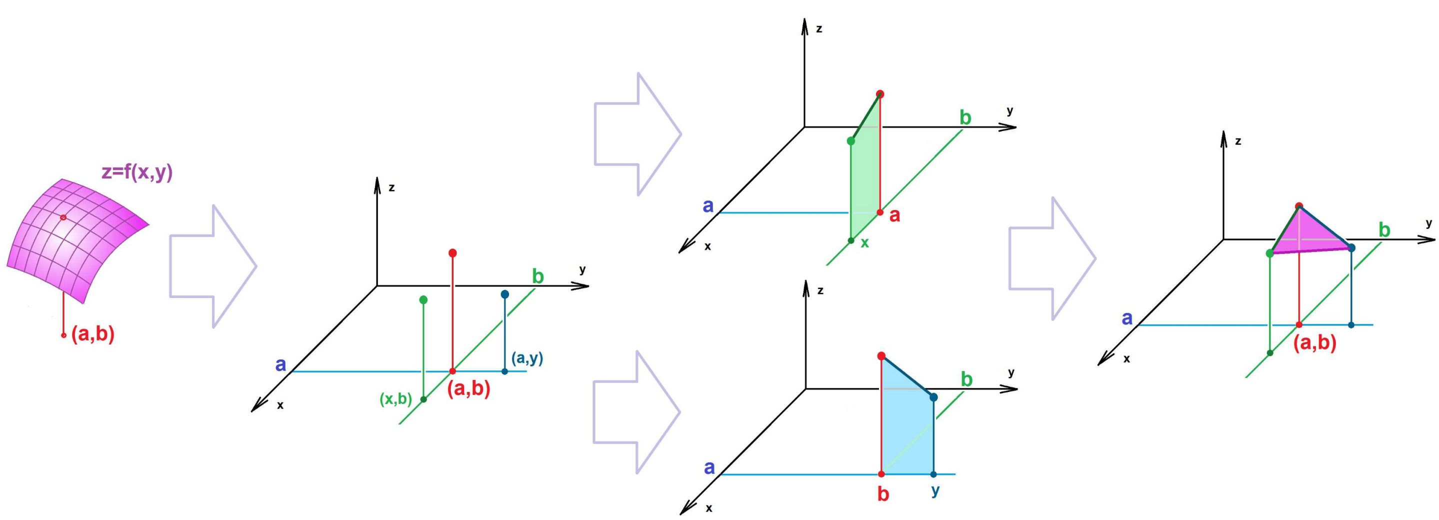
Now, we are interested in the *rate* of change, the change of z relative to the change of X. It has been the ratio of the former over the latter. However, ΔX is a vector this time. What should be the denominator o this fraction?

We choose the increment of X of a specific kind. We only consider the increment of X in one of those directions:

$$\begin{array}{ccccccccc} \Delta X_k & = & < & 0, & ..., & 0, & \Delta x_k, & 0, & ..., & 0 & > \\ \hline & & & 1 & ... & k-1 & k & k+1 & ... & n & \end{array}$$

The partial difference of $z = f(X) = f(x_1, ..., x_n)$ with respect x_k at C is the difference of f with respect to x_k:

$$f(X + \Delta X_k) - f(X)\,.$$



We collect these rates of change into one function, Δf.

Now we just divide by the respective increment of X:

Definition 4.3.2: partial difference quotient

The *partial difference quotient of $z = f(X) = f(x_1, ..., x_n)$ with respect x_k at X* is defined to be the rate of change of z with respect to x_k denoted as follows:

$$\boxed{\frac{\Delta f}{\Delta x_k}(C) = \frac{f(X + \Delta X_k) - f(X)}{\Delta x_k}}$$

where C is the secondary node of the edge between X and $X + \Delta X_k$.

We collect these rates of change into one function!

Definition 4.3.3: difference quotient

The *difference quotient* of $z = f(X) = f(x_1, ..., x_n)$ at edge C is equal to the partial difference quotient of $z = f(X)$ with respect x_k at C denoted as follows:

$$\boxed{\frac{\Delta f}{\Delta X}(C) = \frac{\Delta f}{\Delta x_k}(C)}$$

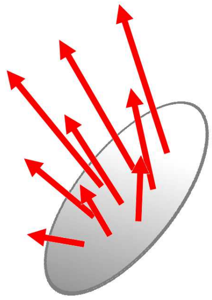
where C is the secondary node of the edge between X and $X + \Delta X_k$.

This is a 1-form in $\mathbf{R}^n$.

Can we see this quantity as a vector? For every edge C aligned with the kth axis, we consider:

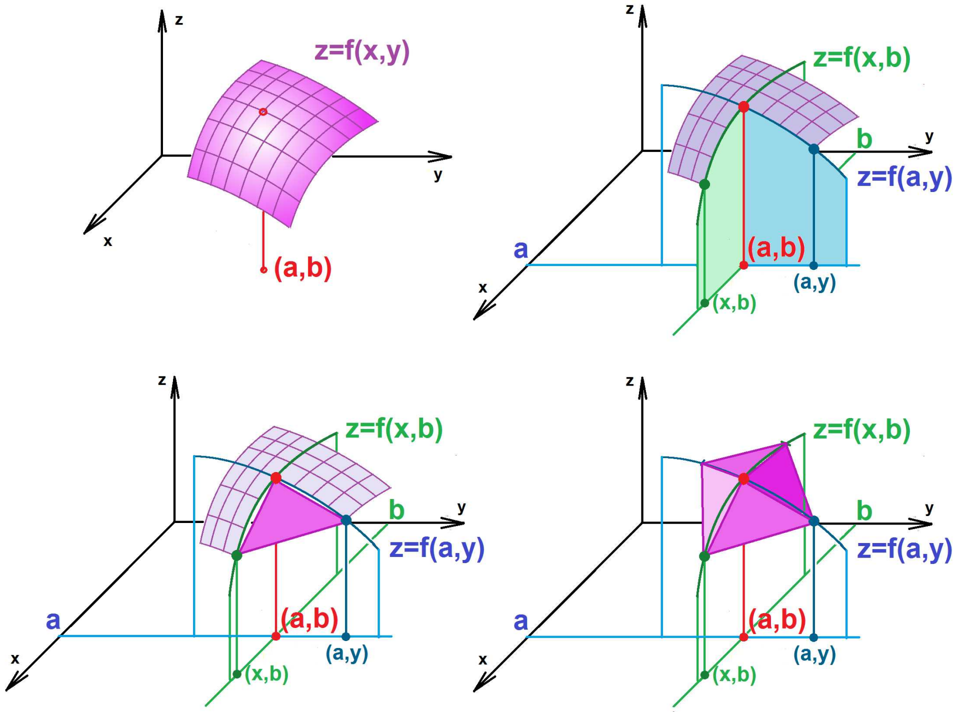
$$\boxed{\left\langle 0, ..., 0, \frac{\Delta f}{\Delta x_k}(C), 0, ..., 0 \right\rangle}$$

The reason is that there is no flow in an direction other than x_k. However, as we move from the discrete case to the continuous, we also move from the hydraulic analogy to *flow on surface*:



4.4. The gradient

For the continuous case, we repeat the construction in the last section (below) and then just take the limit:



We focus on one point $A = (a_1, ..., a_n)$ in $\mathbf{R}^n$.

Definition 4.4.1: partial derivative

The *partial derivative of* $z = f(X) = f(x_1, ..., x_n)$ *with respect* x_k *at* $X = A = (a_1, ..., a_n)$ are defined to be the limit of the partial difference quotient with respect to x_k at $x_k = a_k$, if it exists, denoted as follows:

$$\boxed{\frac{\partial f}{\partial x_k}(A) = \lim_{\Delta x_k \to 0} \frac{\Delta f}{\Delta x_k}(A)}$$

or

$$\boxed{f'_k(A)}$$

The following is an alternative definition:

Theorem 4.4.2: Partial Derivatives

The partial derivative of $z = f(X)$ *with respect to* x_k *at* $X = A = (a_1, ..., a_n)$ *is found as the derivative of the numerical function*

$$g(x) = f(a_1, ..., a_{k-1}, x, a_{k+1}, ..., a_n),$$

evaluated at $x = a_k$*; i.e.,*

$$\boxed{\frac{\partial f}{\partial x_k}(A) = \frac{d}{dx} f\left(a_1, ..., a_{k-1}, x, a_{k+1}, ..., a_n\right)\Big|_{x=a_k}}$$

So, this is the derivative of $z = f(x_1, ..., x_n)$ with respect to x_k with the rest of the variables fixed. These are, of course, just the slopes these edges of the graph.

There is another concept that reappears in dimension n:

Definition 4.4.3: gradient

The *gradient*, or the derivative, of f at $X = A$ is defined to be the limit of the difference quotient as well as the vector of partial derivatives denoted as follows:

$$\boxed{\nabla f(A) = \frac{df}{dX}(A) = \left\langle \frac{\partial f}{\partial x_1}(A), ..., \frac{\partial f}{\partial x_n}(A) \right\rangle}$$

Warning!

The gradient is not the limit of the difference quotient.

Warning!

The gradient notation is to be read as follows:

$$\big(\nabla f\big)(A),\ \big(\operatorname{grad} f\big)(A\,,$$

i.e., the gradient is computed and then evaluated at $X = A$.

Definition 4.4.4: best linear approximation

Suppose $z = f(X)$ is defined at $X = A$ and

$$l(X) = f(A) + M \cdot (X - A)$$

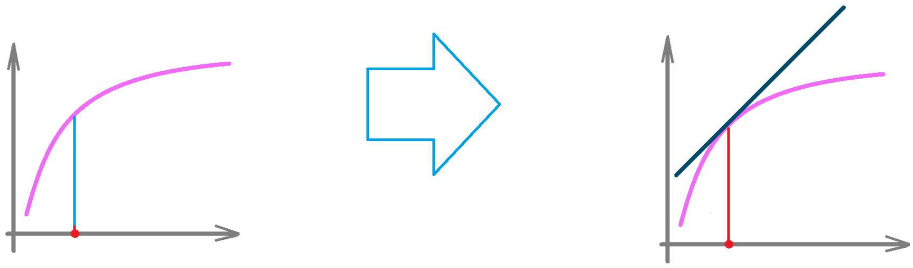
is any of its linear approximations at that point. Then, $z = l(X)$ is called the *best linear approximation* of f at $X = A$ if the following is satisfied:

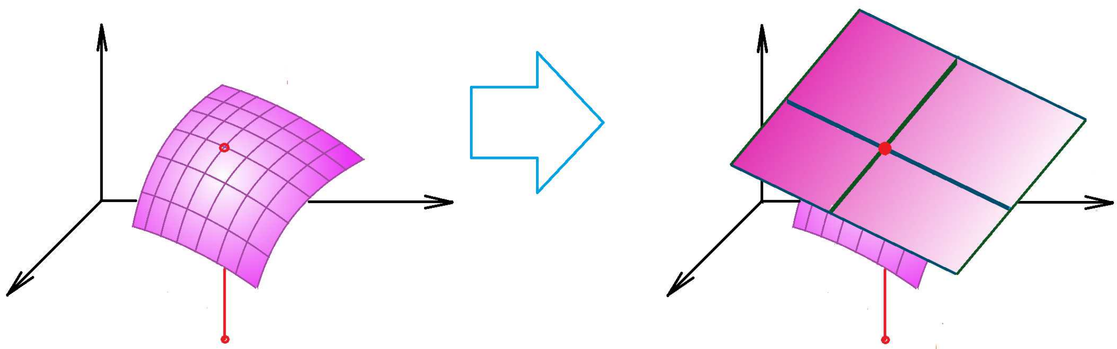
$$\lim_{X \to A} \frac{f(X) - l(X)}{||X - A||} = 0\,.$$

In that case, the function f is called *differentiable* at $X = A$.

The numerator in the formula is the error of the approximation and the denominator is the length of the "run".

The result is the tangent line in dimension 1 and the tangent plane in dimension 2:

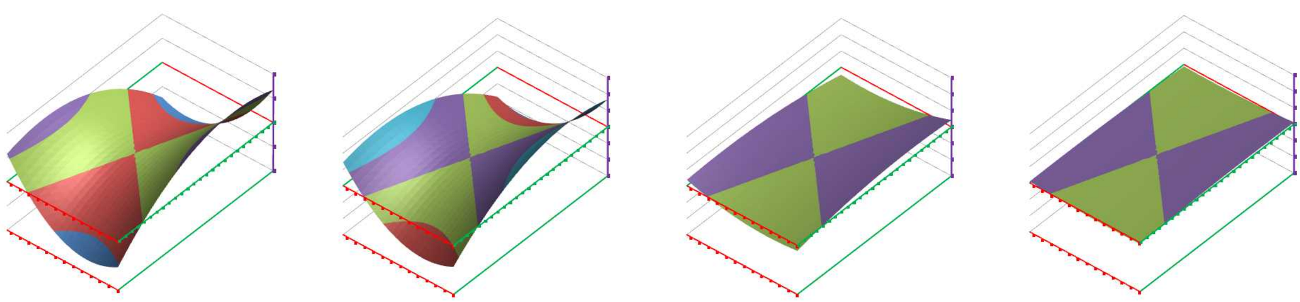




Example 4.4.5: differentiable

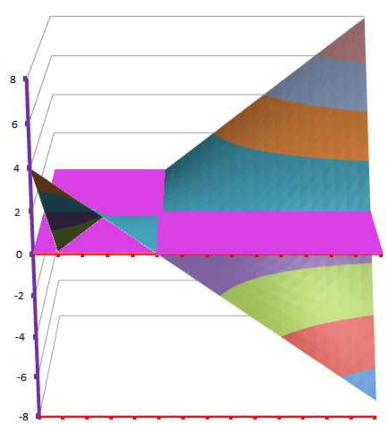
So, we stick with the functions of n variables the graphs of which – on a small scale – look like lines in dimension $n = 1$, like planes in dimension $n = 2$, and generally like $\mathbf{R}^n$!

Below is a visualization of a differentiable function of two variables and its level curves:



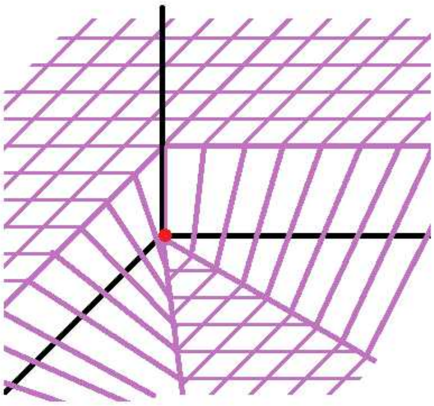
Not only these curves look like straight lines when we zoom in, they also progress at a uniform rate.

Just as the tangent line doesn't have to have a single point in common with the graph, same applies to the tangent plane:

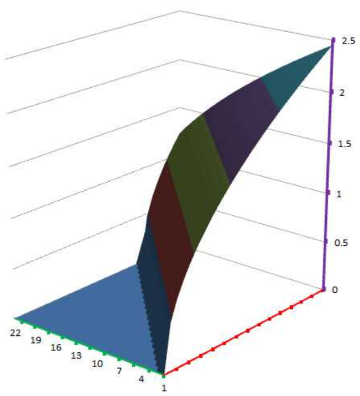


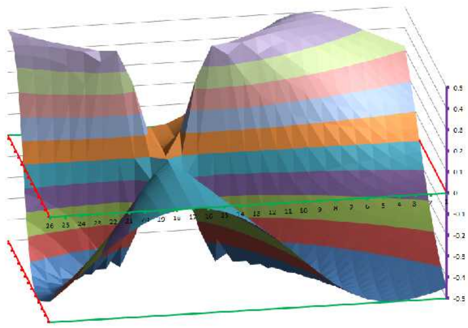
Example 4.4.6: non-differentiable

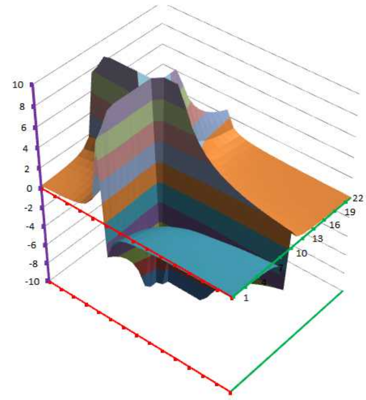
The simple function below is continuous and its partial derivatives are both zero at the origin:



However, because of the cliff, there is no plane that approximates the surface. The following function is also not differentiable:







Theorem 4.4.7: Uniqueness of Best Linear Approximation

For a function differentiable as $X = A$, there is only one best linear approximation at A.

Proof.

By contradiction. Suppose we have two such functions

$$l(X) = f(A) + M \cdot (X - A) \quad \text{with} \quad \lim_{X \to A} \frac{f(X) - l(X)}{||X - A||} = 0\,,$$

and

$$p(X) = f(A) + Q \cdot (X - A) \quad \text{with} \quad \lim_{X \to A} \frac{f(X) - q(X)}{||X - A||} = 0\,.$$

Then we have:

$$\lim_{X \to A} \left(\frac{f(X) - f(A)}{||X - A||} + \frac{M \cdot (X - A)}{||X - A||} \right) = 0\,,$$

and

$$\lim_{X \to A} \left(\frac{f(X) - f(A)}{||X - A||} + \frac{Q \cdot (X - A)}{||X - A||} \right) = 0\,.$$

Therefore, by the *Sum Rule* we have:

$$\lim_{X \to A} \left(\frac{M \cdot (X - A)}{||X - A||} - \frac{Q \cdot (X - A)}{||X - A||} \right) = 0\,,$$

or

$$\lim_{X \to A} \frac{M \cdot (X - A) - Q \cdot (X - A)}{||X - A||} = 0\,,$$

or

$$\lim_{X \to A} \frac{(M - Q) \cdot (X - A)}{||X - A||} = 0\,,$$

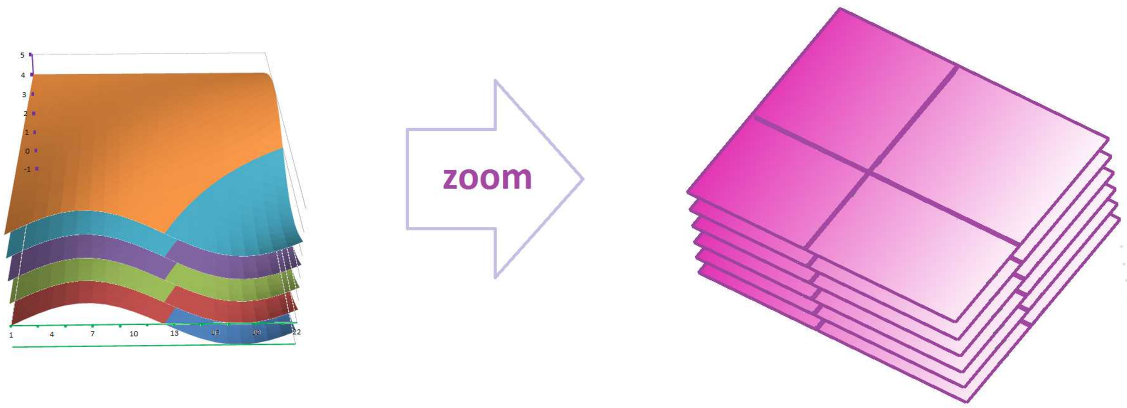
or

$$\lim_{X \to A} (M - Q) \cdot \frac{X - A}{||X - A||} = 0\,.$$

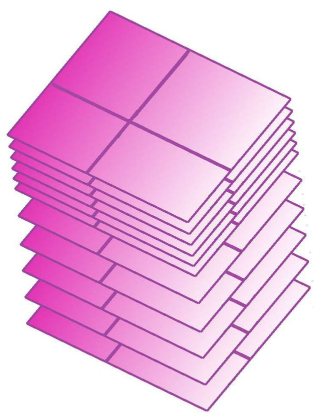
The limit of the fraction, however, does not exist. Therefore, $M - Q = 0$.

Example 4.4.8: 3 variables

Below is a visualization of a differentiable function of three variables given by its level surfaces:



Not only these surfaces look like planes when we zoom in, they also progress at a uniform rate. For example, this function is not differentiable:



The problem of finding the best linear approximation has been solved:

Theorem 4.4.9: Best Linear Approximation

If

$$l(X) = f(A) + M \cdot (X - A)$$

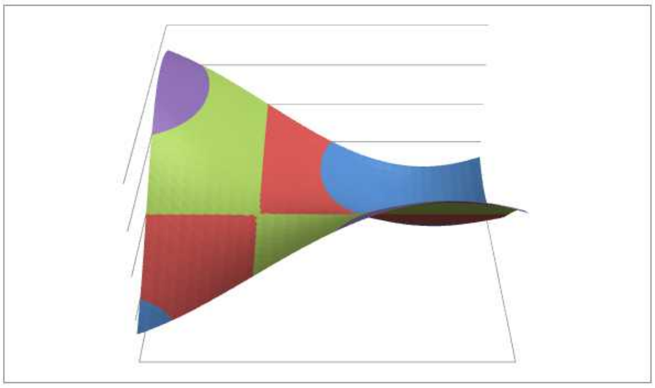
is the best linear approximation of $z = f(X)$ at $X = A$, then

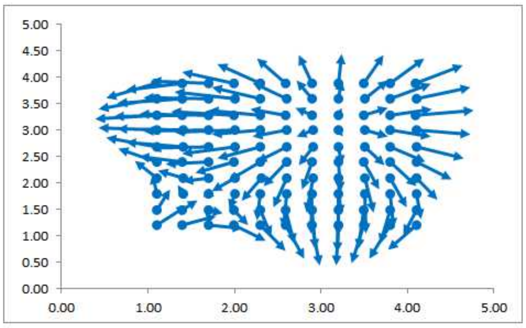
$$\boxed{M = \operatorname{grad} f(A)}$$

Exercise 4.4.10

Prove the theorem.

Now, suppose we carry out this procedure of linear approximation at each location throughout the domain of the function. We will have a vector at each location. This is a *vector field*!





The gradient serves as *the* derivative of a differentiable function.

Warning!

When the function is not differentiable, combining its variables, say x and y, into one, $X = (x, y)$, may be ill-advised even when the partial derivatives make sense.

4.5. Algebraic properties of the difference quotients and the gradients

Just as in dimension 1, differentiation is a special kind of function too, a *function of functions*:

$$f \to \boxed{\frac{d}{dX}} \to G = \nabla f$$

The main difference is that the domain and the range of this function are different. We need to understand how this function operates.

Warning!

Even though the derivative of a parametric curve in $\mathbf{R}^n$ at a point and the derivative of a function of n variables at a point are both vectors in $\mathbf{R}^n$, this doesn't make the two derivatives similar.

We start with *linear functions*. After all, they serve as good-enough substitutes for the functions around a fixed point. This is the "Linear Sum Rule":

	linear function	its gradient
$f(X)$	$= p + M(X - A)$	M
$+$		
$g(X)$	$= q + N(X - A)$	N
$f(X) + g(X)$	$= p + (M + N)(X - A)$	$M + N$

We used the *Linearity of the dot product.* The "Linear Constant Multiple Rule" relies on the same property:

	linear function	its gradient
$f(X)$	$= p + M(X - A)$	M
$\cdot k$		
$k \cdot f(X)$	$= kp + (kM)(X - A)$	kM

The first operation on on functions is addition.

For any two functions of several variables f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the differences defined at the corresponding secondary node C satisfy:

$$\boxed{\Delta(f+g)\,(C) = \Delta f\,(C) + \Delta g\,(C)}$$

The wordings of the actual theorems are identical to those for numerical functions presented in Volume 2:

Theorem 4.5.1: Sum Rule for Differences

The difference of the sum of two functions is the sum of their differences.

In other words, for any two functions f, g, their differences satisfy:

$$\boxed{\Delta(f+g) = \Delta f + \Delta g}$$

Proof.

Applying the definition to the function $f + g$, we have:

$$\begin{aligned}
\Delta(f+g)(C) &= (f+g)(X+\Delta X) - (f+g)(X) \\
&= f(X+\Delta X) + g(X+\Delta X) - f(X) - g(X) \\
&= \big(f(X+\Delta X) - f(X)\big) + \big(g(X+\Delta X) - g(X)\big) \\
&= \Delta f(C) + \Delta g(C)\,.
\end{aligned}$$

The formula holds for every ΔX. We choose it to follow the kth axis and then divide the last formula by Δx_k:

$$\frac{\Delta(f+g)}{\Delta x_k}(C) = \frac{\Delta f}{\Delta x_k}(C) + \frac{\Delta g}{\Delta x_k}(C)\,.$$

These are the n components of the vector of the difference quotient. We combine them into a single formula. For any two functions of several variables f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the difference quotients defined at the corresponding secondary node C satisfy:

$$\boxed{\frac{\Delta(f+g)}{\Delta X}(C) = \frac{\Delta f}{\Delta X}(C) + \frac{\Delta g}{\Delta X}(C)}$$

We have the analog for difference quotients:

Theorem 4.5.2: Sum Rule for Difference Quotients

The difference quotient of the sum of two functions is the sum of their difference quotients.

In other words, for any two functions f, g defined at the adjacent nodes X and $X + \Delta X$ of a partition, the difference quotients (defined at the corresponding secondary node) satisfy:

$$\boxed{\frac{\Delta(f+g)}{\Delta X} = \frac{\Delta f}{\Delta X} + \frac{\Delta g}{\Delta X}}$$

In a way, we just take the limit of the last formula as

$$||\Delta X|| \to 0\,.$$

For any two functions f, g differentiable at $X = A$, we have:

$$\boxed{\frac{d(f+g)}{dX}(A) = \frac{df}{dX}(A) + \frac{dg}{dX}(A)}$$

So, we have the analog for derivatives:

Theorem 4.5.3: Sum Rule for Derivatives

The sum of two functions differentiable at a point is differentiable at that point and its derivative is equal to the sum of their derivatives.

In other words, for any two functions f, g differentiable at X, we have at X:

$$\boxed{\frac{d(f+g)}{dX} = \frac{df}{dX} + \frac{dg}{dX}}$$

Proof.

Suppose

$$l(X) = f(A) + M \cdot (X - A) \quad \text{and} \quad k(X) = g(A) + N \cdot (X - A)$$

are the best linear approximations at A of f and g respectively. Then, the following is satisfied:

$$\lim_{X \to A} \frac{M \cdot (X - A)}{||X - A||} = 0 \ \text{ and } \ \lim_{X \to A} \frac{N \cdot (X - A)}{||X - A||} = 0\,.$$

We can add the two limit together, as allowed by the *Sum Rule for Limits*, and then manipulate the expression:

$$\begin{array}{ll} 0 & = \lim\limits_{X \to A} \dfrac{M \cdot (X - A)}{||X - A||} + \lim\limits_{X \to A} \dfrac{N \cdot (X - A)}{||X - A||} \\ & = \lim\limits_{X \to A} \dfrac{(M + N) \cdot (X - A)}{||X - A||} \end{array}.$$

According to the definition,

$$l(X) + k(X) = f(A) + g(A) + (M + N) \cdot (X - A)$$

is the best linear approximation of $f + g$.

Exercise 4.5.4

Derive the last theorem from the previous.

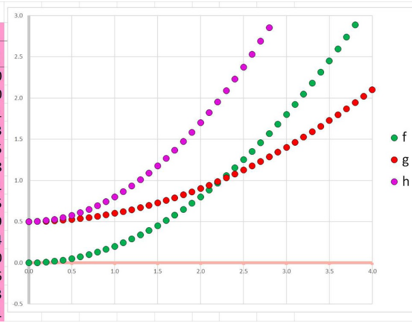
So, Δ's become d's!

In the alternative notation:

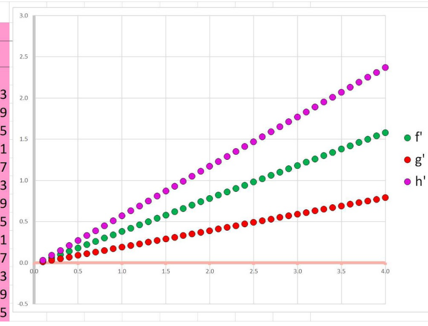
$$\boxed{\nabla(f+g)=\nabla f+\nabla g}$$

Either of the two last theorems can be illustrated with the following:

n=	40				
h=	0.10		1.00	1.00	
		x	f	g	h=f+g
	1	0.0	0.00	0.50	0.50
	2	0.1	0.00	0.50	0.50
	3	0.2	0.01	0.50	0.51
	4	0.3	0.02	0.51	0.53
	5	0.4	0.03	0.52	0.55
	6	0.5	0.05	0.53	0.58
	7	0.6	0.07	0.54	0.61
	8	0.7	0.10	0.55	0.65
	9	0.8	0.13	0.56	0.69
	10	0.9	0.16	0.58	0.74
	11	1.0	0.20	0.60	0.80
	12	1.1	0.24	0.62	0.86
	13	1.2	0.29	0.64	0.93
	14	1.3	0.34	0.67	1.01



n=	40				
h=	0.10				
		x	f'	g'	h'
	1	0.0			
	2	0.1	0.02	0.01	0.03
	3	0.2	0.06	0.03	0.09
	4	0.3	0.10	0.05	0.15
	5	0.4	0.14	0.07	0.21
	6	0.5	0.18	0.09	0.27
	7	0.6	0.22	0.11	0.33
	8	0.7	0.26	0.13	0.39
	9	0.8	0.30	0.15	0.45
	10	0.9	0.34	0.17	0.51
	11	1.0	0.38	0.19	0.57
	12	1.1	0.42	0.21	0.63
	13	1.2	0.46	0.23	0.69
	14	1.3	0.50	0.25	0.75



The second operation on functions is scalar multiplication.

For any real k and any function f defined at the nodes X and $X+\Delta X$ of the partition, we have the differences defined at the corresponding secondary node C satisfy:

$$\boxed{\Delta(k\cdot f)\,(C)=k\cdot\Delta f\,(C)}$$

The result is indistinguishable from the one for numerical functions:

Theorem 4.5.5: Constant Multiple Rule for Differences

The difference of a multiple of a function is the multiple of the function's difference.

In other words, for any function f, the its difference satisfies:

$$\boxed{\Delta(kf)=k\Delta f}$$

Proof.

Applying the definition to the function $k\,f$, we have:

$$\begin{aligned}\Delta(k\cdot f)(C)&=(k\cdot f)(X+\Delta X)-(k\cdot f)(X)\\&=k\cdot f(X+\Delta X)-k\cdot f(X)\\&=k\cdot\big(f(X+\Delta X)-f(X)\big)\\&=k\cdot\Delta f\,(C)\,.\end{aligned}$$

The formula holds for every ΔX. We choose it to follow the kth axis and then divide the last formula by Δx_k:

$$\frac{\Delta(k\cdot f)}{\Delta x_k}(C)=k\cdot\frac{\Delta f}{\Delta x_k}(C)\,.$$

These are the n components of the vector of the difference quotient. We combine them into a single formula. For any real k and any function f defined at the nodes X and $X+\Delta X$ of the partition, we have the difference quotients defined at the corresponding secondary node C satisfy:

$$\boxed{\frac{\Delta(k\cdot f)}{\Delta X}(C)=k\cdot\frac{\Delta f}{\Delta X}(C)}$$

We have the analog for difference quotients:

Theorem 4.5.6: Constant Multiple Rule for Difference Quotients

The difference quotient of a multiple of a function is the multiple of the function's difference quotient.

In other words, for any function f defined at the adjacent nodes X and $X+\Delta X$ of a partition and any real k, the difference quotients (defined at the corresponding secondary node) satisfy:

$$\boxed{\frac{\Delta(kf)}{\Delta X} = k\frac{\Delta f}{\Delta X}}$$

Again, we take the limit of the last formula as $||\Delta X|| \to 0$. For any real k and any function f differentiable at $X = A$, we have:

$$\boxed{\frac{d(k \cdot f)}{dX}(A) = k \cdot \frac{df}{dX}(A)}$$

We have the analog for derivatives:

Theorem 4.5.7: Constant Multiple Rule for Derivatives

A multiple of a function differentiable at a point is differentiable at that point, and its derivative is equal to the multiple of the function's derivative.

In other words, for any function f differentiable at x and any real k, we have at X:

$$\boxed{\frac{d(kf)}{dX} = k\frac{df}{dX}}$$

Exercise 4.5.8

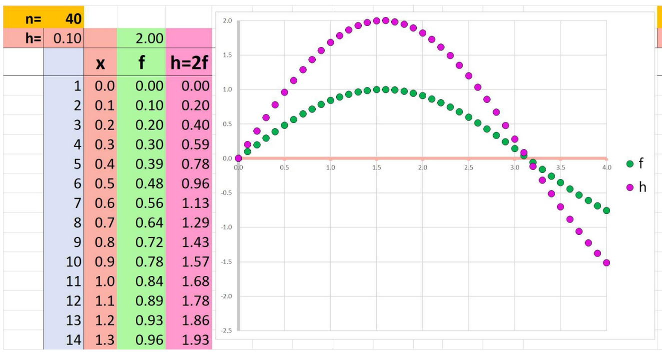
Derive the last theorem from the previous.

And Δ's become d's again.

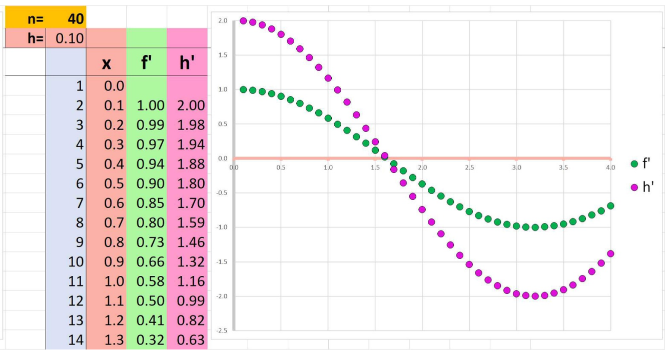
In the alternative notation:

$$\boxed{\nabla(kf) = k\nabla f}$$

Either of the two last theorems can be illustrated with the following:

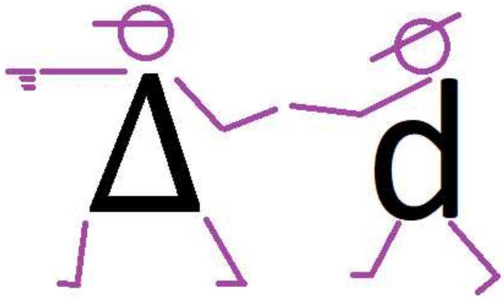


n=	40			
h=	0.10		2.00	
		x	f	h=2f
	1	0.0	0.00	0.00
	2	0.1	0.10	0.20
	3	0.2	0.20	0.40
	4	0.3	0.30	0.59
	5	0.4	0.39	0.78
	6	0.5	0.48	0.96
	7	0.6	0.56	1.13
	8	0.7	0.64	1.29
	9	0.8	0.72	1.43
	10	0.9	0.78	1.57
	11	1.0	0.84	1.68
	12	1.1	0.89	1.78
	13	1.2	0.93	1.86
	14	1.3	0.96	1.93



n=	40			
h=	0.10			
		x	f'	h'
	1	0.0		
	2	0.1	1.00	2.00
	3	0.2	0.99	1.98
	4	0.3	0.97	1.94
	5	0.4	0.94	1.88
	6	0.5	0.90	1.80
	7	0.6	0.85	1.70
	8	0.7	0.80	1.59
	9	0.8	0.73	1.46
	10	0.9	0.66	1.32
	11	1.0	0.58	1.16
	12	1.1	0.50	0.99
	13	1.2	0.41	0.82
	14	1.3	0.32	0.63

As we see (and will see again in this chapter), the derivative follows the difference and the difference quotient, every time:

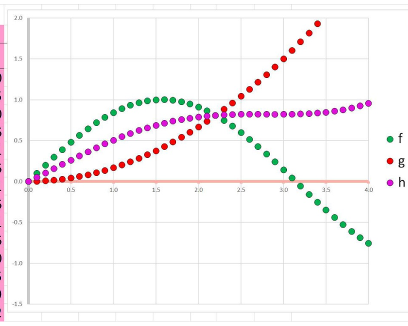


These two operations can be combined into one producing *linear combinations*:

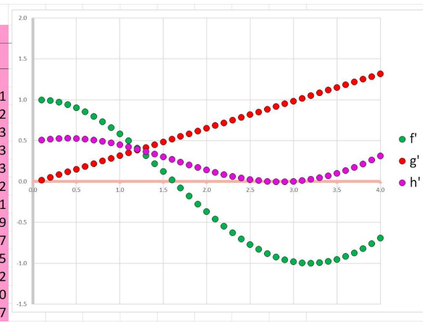
$$\alpha x + \beta y\,,$$

where α, β are two constant numbers. The idea applies to functions; for example, this is the average of two functions (left):

n=	40				
h=	0.10		0.50	0.50	
		x	f	g	h
	1	0.0	0.00	0.00	0.00
	2	0.1	0.10	0.00	0.05
	3	0.2	0.20	0.01	0.10
	4	0.3	0.30	0.02	0.16
	5	0.4	0.39	0.03	0.21
	6	0.5	0.48	0.04	0.26
	7	0.6	0.56	0.06	0.31
	8	0.7	0.64	0.08	0.36
	9	0.8	0.72	0.11	0.41
	10	0.9	0.78	0.14	0.46
	11	1.0	0.84	0.17	0.50
	12	1.1	0.89	0.20	0.55
	13	1.2	0.93	0.24	0.59
	14	1.3	0.96	0.28	0.62



n=	40				
h=	0.10		0.50	0.50	
		x	f'	g'	h'
	1	0.0			
	2	0.1	1.00	0.02	0.51
	3	0.2	0.99	0.05	0.52
	4	0.3	0.97	0.08	0.53
	5	0.4	0.94	0.12	0.53
	6	0.5	0.90	0.15	0.53
	7	0.6	0.85	0.18	0.52
	8	0.7	0.80	0.22	0.51
	9	0.8	0.73	0.25	0.49
	10	0.9	0.66	0.28	0.47
	11	1.0	0.58	0.32	0.45
	12	1.1	0.50	0.35	0.42
	13	1.2	0.41	0.38	0.40
	14	1.3	0.32	0.42	0.37



We also notice what happens to their derivatives (right):

▶ The derivative of the average is the average of the derivatives.

The question becomes: What happens to linear combinations of functions under differentiation?

Recall that a function F is linear if it "preserves" linear combinations:

$$\alpha x + \beta y \ \to \ \boxed{\quad F \quad} \ \to \ \alpha F(x) + \beta F(y)$$

With this idea, these two formulas can be combined into one: The difference, the difference quotient, and the derivative are linear functions of functions. A precise version is below:

Theorem 4.5.9: Linearity of Differentiation

The difference, the difference quotient, and the derivative of a linear combination of two functions of several veraibles is the linear combination of their differences, difference quotients, and derivatives respectively, whenever they exist.

In other words, we have:

$$\boxed{\begin{array}{lll}
\Delta(\alpha f + \beta g) & = \alpha \Delta f & + \beta \Delta g \\
\dfrac{\Delta(\alpha f + \beta g)}{\Delta X} & = \alpha \dfrac{\Delta f}{\Delta X} & + \beta \dfrac{\Delta g}{\Delta X} \\
\dfrac{d(\alpha f + \beta g)}{dX} & = \alpha \dfrac{df}{dX} & + \beta \dfrac{dg}{dX}
\end{array}}$$

The last formula is illustrated with the following diagram:

$$\alpha f + \beta g \ \to \ \boxed{\dfrac{d}{dX}} \ \to \ \alpha \nabla f + \beta \nabla g$$

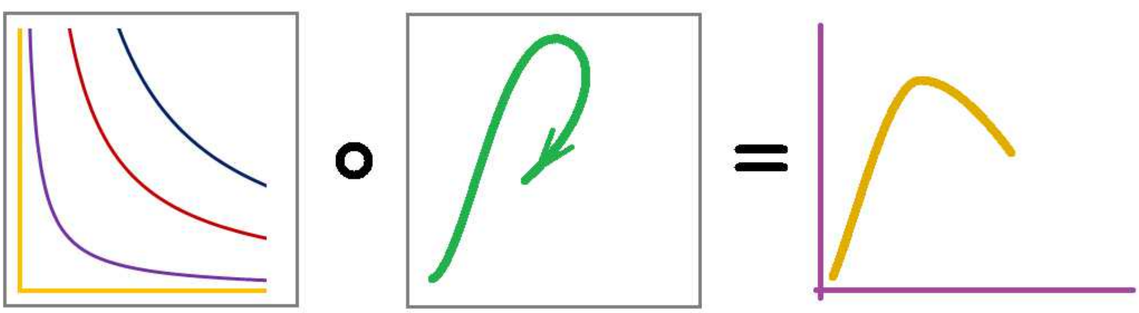
4.6. Compositions and the Chain Rule

How does one learn the *terrain* around him without the ability to fly? By taking *hikes* around the area! Mathematically, the former is a function of two variables and the latter is a parametric curve. Furthermore, we examine the surface of the graph of this function via its *composition* with these parametric curves.

There are two functions with which to compose a function of several variables: a parametric curve before or a numerical function after. This is the former:

trip map				
t	$\longrightarrow$	(x,y)	$\longrightarrow$	z
		terrain map		

Recall how we interpret this composition. We imagine creating a *trip plan* as a parametric curve $X=F(t)$: the times and the places put on a simple automotive map, and then bring the *terrain map* of the area as a function of two variables $z=f(x,y)$:

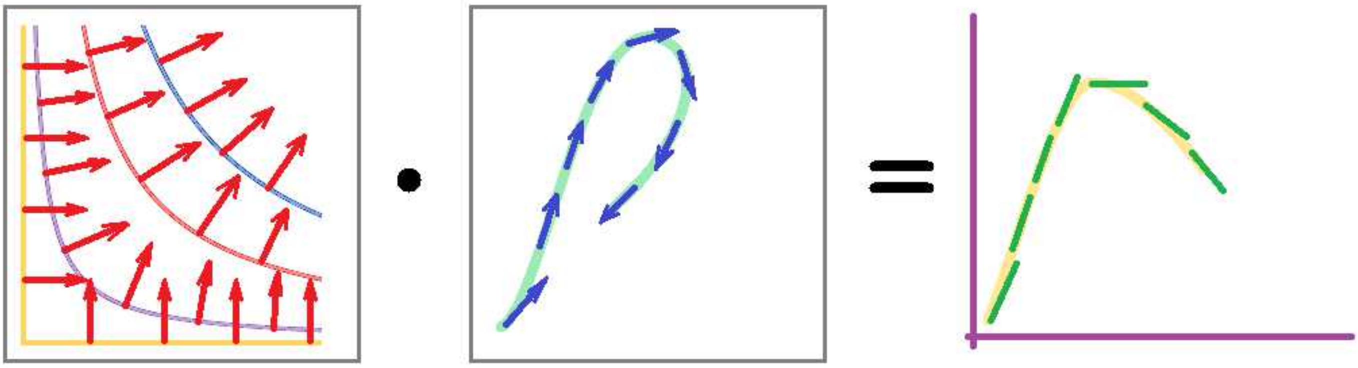


The former give us the location for every moment of time and the latter the elevation for every location. Their composition gives us the elevation for every moment of time.

To understand how the derivatives of these two functions are combined, we start with *linear functions*. In other words, what if we travel along a straight line on a flat, not necessarily horizontal, surface (maybe a roof)? After this simple substitution, the derivatives are found by direct examination:

		linear function	its derivative	
parametric curve:	$X=F(t)$	$=A+D(t-a)$	D	in $\mathbf{R}^n$
	$\circ$			
function of several variables:	$z=f(X)$	$=p+M\cdot(X-A)$	M	in $\mathbf{R}^n$
numerical function:	$f(F(t))$	$=p+M\cdot(A+D(t-a)-A)$		
		$=p+(M\cdot D)(t-a)$	$M\cdot D$	in $\mathbf{R}$

Thus, *the derivative of the composition is the dot product of the two derivatives.*



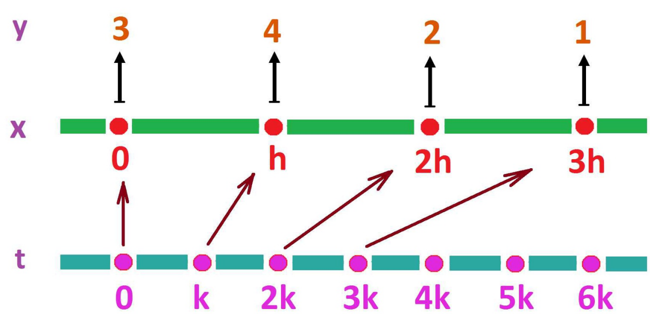
We use this result for the general case of arbitrary differentiable functions via their linear approximations. The result is understood in the same way as in dimension 1:

- If we double our horizontal speed (with the same terrain), the climb will be twice as fast.
- If we double steepness of the terrain (with the horizontal speed), the climb will be twice as fast.

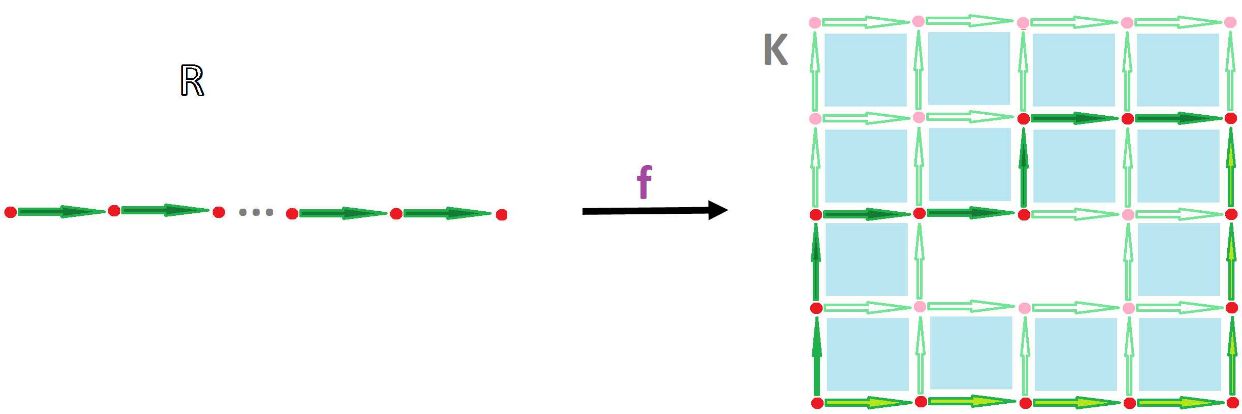
It follows that the speed of the climb is proportional to both our horizontal speed and the steepness of the terrain. This number is computed as the dot product of:

- The derivative of the parametric curve F of the trip, i.e., the horizontal velocity $\left\langle \frac{dx}{dt}, \frac{dy}{dt} \right\rangle$, and
- The gradient of the terrain function f, i.e., $\left\langle \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \right\rangle$.

For the discrete case, we need the parametric curve $X = F(t)$ to map the partition for t to the partition for X. In other words, it has to follow the grid. Dimension 1:



Dimension 2:



The dependence of the variables is as follows:

$$t \xrightarrow{F} X \xrightarrow{g} z$$

Both t and X have partitions.

Recall that we have noticed this pattern of cancellation:

$$\frac{\Delta X}{\Delta t} \cdot \frac{\Delta z}{\Delta X} = \frac{\Delta z}{\Delta t}$$

Theorem 4.6.1: Chain Rule for Differences

The difference of the composition of two functions is the difference of the second function.

In other words, for any function $X = F(t)$ defined at two adjacent nodes t and $t + \Delta t$ of a partition, and any function $z = g(X)$ defined at the two adjacent nodes $X = F(t)$ and $X + \Delta X = F(t + \Delta t)$ of a partition, we have the difference quotients (defined at the secondary nodes c and $Q = F(c)$ within these edges of the two partitions respectively) satisfy:

$$\boxed{\Delta(g \circ F)(C) = \Delta g(Q)}$$

We have the analog for the difference quotients:

Theorem 4.6.2: Chain Rule for Difference Quotients

The difference quotient of the composition of two functions is equal to the product of the two difference quotients.

In other words, for any function $X = F(t)$ defined at two adjacent nodes t and $t + \Delta t$ of a partition, and any function $z = g(X)$ defined at the two adjacent nodes $X = F(t)$ and $X + \Delta X = F(t + \Delta t)$ of a partition, we have the difference quotients (defined at the secondary nodes c and $Q = F(c)$ within these edges of the two partitions respectively) satisfy, provided $\Delta X \neq 0$:

$$\boxed{\frac{\Delta(g \circ F)}{\Delta t}(c) = \frac{\Delta g}{\Delta X}(Q) \cdot \frac{\Delta F}{\Delta t}(c)}$$

Proof.

Suppose F moves along the grid at $x = c$. Then

$$\Delta F(c) = \Delta X_k =< 0, ..., 0, \Delta x_k, 0, ..., 0 > ,$$

for some k and $\Delta x_k \neq 0$. Then to get to the difference quotients, we just multiply the formula for the differences on the left and on the right by these respectively:

$$\frac{1}{\Delta t} = \frac{1}{\Delta x_k} \cdot \frac{\Delta x_k}{\Delta t} .$$

As a result, the partial difference quotient of g emerges:

$$\frac{\Delta(g \circ F)}{\Delta t}(c) = \frac{\Delta g(Q)}{\Delta x_k} \cdot \frac{\Delta x_k}{\Delta t} = \frac{\Delta g}{\Delta X}(Q) \cdot \frac{\Delta F}{\Delta t}(c) .$$

Alternatively, because the increment of X is very simple, so is the difference quotient of F:

$$\frac{\Delta F}{\Delta t}(c) = \frac{1}{\Delta t}\Delta X_k = \left\langle 0, ..., 0, \frac{\Delta x_k}{\Delta t}, 0, ..., 0 \right\rangle .$$

The dot product emerges:

$$\left\langle ..., \frac{\Delta g(Q)}{\Delta x_k}, ... \right\rangle \cdot \left\langle 0, ..., 0, \frac{\Delta x_k}{\Delta t}, 0, ..., 0 \right\rangle = \frac{\Delta g(Q)}{\Delta x_k} \frac{\Delta x_k}{\Delta t} .$$

The limit gives us the derivative:

Theorem 4.6.3: Chain Rule for Derivatives

The composition of a function differentiable at a point and a function differentiable at the value of that point under the first function is differentiable at that point, and its derivative is equal to the product of the two derivatives.

In other words, if $X = F(t)$ is differentiable at $t = c$ and $z = g(X)$ is differentiable at $X = Q = F(c)$, then we have:

$$\boxed{\frac{d(g \circ F)}{dt}(c) = \frac{dg}{dX}(Q) \cdot \frac{dF}{dt}(c)}$$

Warning!

We have a dot product on the right.

Exercise 4.6.4

Derive the last theorem from the previous.

Exercise 4.6.5

Find another, non-constant, example of a function $X = F(t)$ such that ΔF may be zero even for small values of Δt.

Warning!

There are two input variables in the right-hand side of each formula; they are linked by substitution.

We see the same pattern of cancellation:

$$\frac{\color{red}{dX}}{dt} \cdot \frac{dz}{\color{red}{dX}} = \frac{dz}{dt}$$

These aren't fractions though.

The formula in the Lagrange notation is as follows:

$$\boxed{(g \circ F)'(t) = \nabla g(F(t)) \cdot F'(t)}$$

Without the input variable:

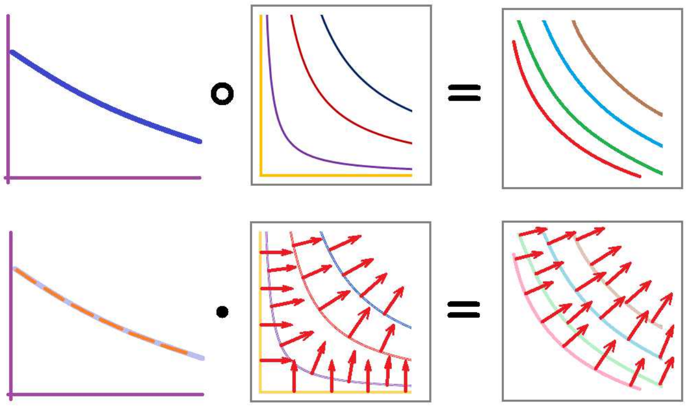
$$\boxed{(g \circ F)' = (\nabla g \circ F) \cdot F'}$$

A function of several variables may appear in another context...

This is the meaning of the composition when our function of several variables is followed by a numerical function:

$$\begin{array}{|ccccc|}
\hline
\multicolumn{4}{|c|}{\text{terrain map}} & \\
\hline
(x,y) & \longrightarrow & z & \longrightarrow & u \\
\hline
& & \multicolumn{3}{|c|}{\text{pressure}} \\
\hline
\end{array}$$

Recall how we interpret this composition. In addition to the *terrain map* of the area as a function of two variables $z = f(x, y)$, we have the *atmospheric pressure* dependent on the elevation (above the sea level) as a numerical function:



The former give us the elevation for every location and the latter the pressure for every elevation. Their composition gives us the pressure for every location.

To understand how the derivatives of these two functions are combined, we start with *linear functions*. After this simple substitution, the derivatives are found by direct examination:

		linear function	its derivative	
function of several variables:	$z = f(X)$	$= p + M \cdot (X - A)$	M	in $\mathbf{R}^n$
	$\circ$			
numerical function:	$u = g(z)$	$= q + m(z - p)$	m	in $\mathbf{R}$
function of several variables:	$g(f(X))$	$= q + m(p + M \cdot (X - A) - p)$		
		$= q + (mM) \cdot (X - A))$	mM	in $\mathbf{R}^n$

Thus, *the derivative of the composition is the scalar product of the two derivatives.*

Once again, the parametric curve $z = f(X)$ has to map the partition for X to the partition for z.

Theorem 4.6.6: Chain Rule For Differences II

The difference of the composition of two functions is found as the difference of the latter; i.e., for any function of several variables $z = f(X)$ defined at adjacent nodes X and $X + \Delta X$ of a partition and any numerical function $u = g(z)$ defined at the adjacent nodes $z = f(X)$ and $z + \Delta z = f(X + \Delta X)$ of a partition, we have the differences (defined at the secondary nodes A and $a = f(A)$ within these edges of the two partitions respectively) satisfy:

$$\boxed{\Delta(g \circ f)(A) = \Delta g\,(a)}$$

Theorem 4.6.7: Chain Rule For Difference Quotients II

The difference quotient of the composition of two functions is found as the product of the two difference quotients; i.e., for any function of several variables $z = f(X)$ defined at adjacent nodes X and $X + \Delta X$ of a partition and any numerical function $u = g(z)$ defined at the adjacent nodes $z = f(X)$ and $z + \Delta z = f(X + \Delta X)$ of a partition, we have the difference quotients (defined at the secondary nodes A and $a = f(A)$ within these edges of the two partitions

respectively) satisfy:

$$\boxed{\frac{\Delta(g\circ f)}{\Delta X}(A)=\frac{\Delta g}{\Delta z}(a)\cdot\frac{\Delta f}{\Delta X}(A)}$$

Note: While the right-hand side above involves a dot product, the one below is a scalar product.

Theorem 4.6.8: Chain Rule For Derivatives II

The composition of a function differentiable at a point and a function differentiable at the image of that point is differentiable at that point and its derivative is found as the product of the two derivatives. In other words, if a function of several variables $z=f(X)$ is differentiable at $X=A$ and a numerical function $u=g(z)$ is differentiable at $a=f(A)$, then we have:

$$\boxed{\frac{d(g\circ f)}{dX}(A)=\frac{dg}{dz}(a)\cdot\frac{df}{dX}(A)}$$

Notice how the intermediate variable is "cancelled" in the Leibniz notation in both of the two forms of the Chain Rule; first:

$$\frac{dz}{\cancel{d}X}\cdot\frac{\cancel{d}X}{dt}=\frac{dz}{dt}\,;$$

and second:

$$\frac{du}{\cancel{d}z}\cdot\frac{\cancel{d}z}{dX}=\frac{du}{dX}\,.$$

Thus, in spite of the fact that these two compositions are very different, the Chain Rule has a somewhat informal – but single – verbal interpretation: *the derivative of the composition of two functions is the product of the two derivatives.* The word "product", as we just saw, is also ambiguous. We saw the multiplication of two numbers in the beginning of the book, then the dot product of two vectors, and finally a vector and a number:

$$\begin{array}{llll}(f\circ g)'(x) & =f'(g(x)) & \cdot & g'(x)\\ (g\circ F)'(x) & =\nabla g(F(t)) & \cdot & F'(t)\\ \nabla(g\circ f)(X) & =g'(f(X)) & \cdot & \nabla f(X)\end{array}$$

The context determines the meaning and this ambiguity serves a purpose: we will see later how this wording is, in a rigorous way, applicable to the composition of *any* two functions.

4.7. Differentiation under multiplication and division

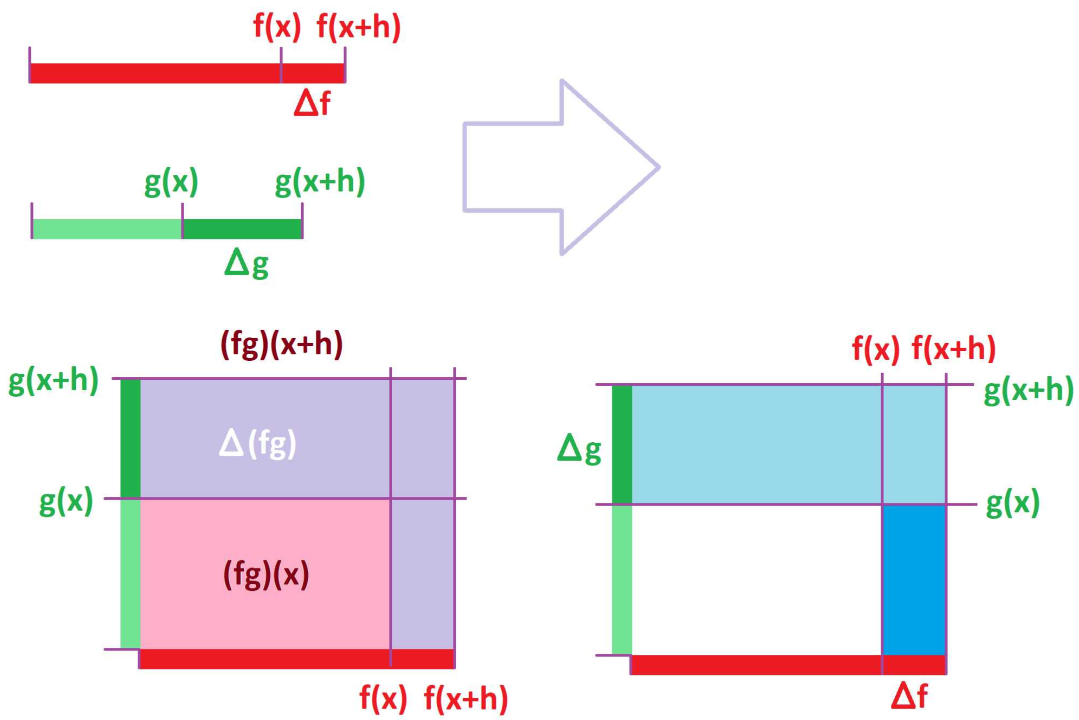
Unfortunately, multiplication and division of linear functions do not produce linear functions...

Warning!

Just as in the case of numerical functions, we face and reject the "naive" product rule: the derivative of the product is not the product of the derivatives! Not only the units don't match, it's worse this time: all three of the derivatives are vectors

and the product of two can't give us the third...

Recall that the product of two *numerical* functions is interpreted as the *areas* of the rectangles formed by the functions (top):



As the width and the depth are increasing, so is the area of the rectangle. We can see that the increase of the area (bottom left) cannot be expressed entirely in terms of the increases of the width and depth! This increase is split into two rectangles (bottom right) corresponding to the two terms in our formula. For any two functions f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the differences defined at the corresponding secondary node C satisfy:

$$\Delta(f \cdot g)\,(C) = f(X + \Delta X) \cdot \Delta g\,(C) + \Delta f\,(C) \cdot g(X)$$

Theorem 4.7.1: Product Rule for Differences

The difference of the product of two functions is found as the sum of the product of the function and the other function's difference.

In other words, for any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the differences (defined at the corresponding secondary node c) satisfy:

$$\Delta(f \cdot g)(C) = f(X + \Delta X) \cdot \Delta g(C) + \Delta f(C) \cdot g(X)$$

Proof.

The trick is to insert extra terms:

$$\begin{aligned}\Delta(f\cdot g)(C) &= (f\cdot g)(X+\Delta X)-(f\cdot g)(X)\\ &= f(X+\Delta X)\cdot g(X+\Delta X)-f(X)\cdot g(X)\\ &= f(X+\Delta X)\cdot g(X+\Delta X)-f(X+\Delta X)\cdot g(X)+f(X+\Delta X)\cdot g(X)-f(X)\cdot g(X)\\ &= f(X+\Delta X)\cdot(g(X+\Delta X)-g(X))+(f(X+\Delta X)-f(X))\cdot g(X)\\ &= f(X+\Delta X)\cdot\Delta g(C)+\Delta f(X)\cdot g(X)\,.\end{aligned}$$

For any two functions f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the difference quotients defined at the corresponding secondary node C satisfy:

$$\frac{\Delta(f\cdot g)}{\Delta X}(C) = f(X+\Delta X)\cdot\frac{\Delta g}{\Delta X}(C)+\frac{\Delta f}{\Delta X}(C)\cdot g(X)$$

So, we just divide by ΔX:

Theorem 4.7.2: Product Rule for Difference Quotients

The difference quotient of the product of two functions is found as the sum of the product of the function and the other function's difference quotient.

In other words, for any two functions f, g defined at the adjacent nodes X and $X + \Delta X$ of a partition, the difference quotients (defined at the corresponding secondary node C) satisfy:

$$\frac{\Delta(f\cdot g)}{\Delta X}(C) = f(X+\Delta X)\cdot\frac{\Delta g}{\Delta x}(C)+\frac{\Delta f}{\Delta x}(C)\cdot g(X)$$

Given two functions f, g differentiable at $X = A$, we have:

$$\frac{d(f\cdot g)}{dX}(A) = f(A)\cdot\frac{dg}{dX}(A)+\frac{df}{dX}(A)\cdot g(A)$$

So, ΔX goes to 0, and all Δs – except for one – are replaced with d's:

Theorem 4.7.3: Product Rule for Derivatives

The product of two functions differentiable at a point is differentiable at that point and its derivative is found as the sum of the product of the function and the other function's derivative.

In other words, for any two functions f, g differentiable at X, we have:

$$\frac{d(f\cdot g)}{dX} = f\cdot\frac{dg}{dX}+\frac{df}{dX}\cdot g$$

Exercise 4.7.4

Derive the last theorem from the previous.

As an informal abbreviation:

▶ When we *multiply* functions, we *cross-multiply* the functions and their derivatives.

In summary, this is how the cross-multiplication works:

	functions	derivatives
first	f	∇f
second	g	∇g

$$\longrightarrow \ \nabla(fg) = f\nabla g + \nabla f g$$

The formula is identical to that for numerical functions but we have to examine it carefully; same things have changed! Indeed, in the right hand side either term is the product of the value of one of the functions (a number) and the value of the gradient of the other (a vector). Furthermore we have a vector at the end of the computation:

$$\underbrace{\underbrace{\overset{\text{scalar}}{f(A)} \cdot \overset{\text{vector}}{\nabla g\,(A)}}_{\text{vector}} + \underbrace{\overset{\text{vector}}{\nabla f\,(A)} \cdot \overset{\text{scalar}}{g(A)}}_{\text{vector}}}_{\text{vector}}$$

It matches the left-hand side.

Moreover, when A varies, the formulas take the form with the algebraic operations discussed in the last section:

$$\nabla(f \cdot g) = f \cdot \nabla g + \nabla f \cdot g\,.$$

Here, either term is the product of one of the functions, a *scalar function*, and the gradient of the other, a *vector field*. Such a product is again a vector field and so is their sum. It matches the left-hand side.

Now division.

For any two functions f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the differences defined at the corresponding secondary node C satisfy:

$$\boxed{\Delta(f/g)\,(C) = \frac{f(X+\Delta X) \cdot \Delta g\,(C) - \Delta f\,(C) \cdot g(X)}{g(X)g(X+\Delta X)}}$$

Theorem 4.7.5: Quotient Rule for Differences

For any two functions f, g defined at the adjacent nodes X and $X + \Delta X$ of a partition, the differences (defined at the corresponding secondary node C) satisfy:

$$\boxed{\Delta(f/g)(C) = \frac{f(X+\Delta X) \cdot \Delta g(C) - \Delta f(C) \cdot g(X)}{g(X)g(X+\Delta X)}}$$

Proof.

We start with the case $f = 1$. Then we have:

$$\begin{aligned} \Delta(1/g)(X) &= \frac{1}{g(X+\Delta X)} - \frac{1}{g(X)} \\ &= \frac{g(X) - g(X+\Delta X)}{g(X+\Delta X)g(X)}\,. \end{aligned}$$

Now the general formula follows from the *Product Rule*.

For any two functions f, g defined at the nodes X and $X + \Delta X$ of the partition, we have the difference quotients defined at the corresponding secondary node C satisfy:

$$\boxed{\frac{\Delta(f/g)}{\Delta X}(C) = \frac{f(X+\Delta X)\cdot \frac{\Delta g}{\Delta X}(C) - \frac{\Delta f}{\Delta X}(C)\cdot g(X)}{g(X)g(X+\Delta X)}}$$

provided $g(X)$, $g(X+\Delta X) \neq 0$.

Theorem 4.7.6: Quotient Rule for Difference Quotients

For any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the difference quotients (defined at the corresponding secondary node c) satisfy:

$$\boxed{\frac{\Delta(f/g)}{\Delta x}(C) = \frac{f(X+\Delta X)\cdot \frac{\Delta g}{\Delta X}(C) - \frac{\Delta f}{\Delta X}(C)\cdot g(X)}{g(X)\cdot g(X+\Delta X)}}$$

provided $g(X)$, $g(X+\Delta X) \neq 0$.

Proof.

We start with the case $f = 1$. Then we have:

$$\begin{aligned}\frac{\Delta(1/g)(X)}{\Delta X} &= \frac{g(X) - g(X+\Delta X)}{\Delta X g(X+\Delta X)g(X)}\\ &= -\frac{g(X+\Delta X)-g(X)}{\Delta X}\cdot\frac{1}{g(X+\Delta X)\cdot g(X)}\\ &= -\frac{\Delta g}{\Delta X}(C)\cdot\frac{1}{g(X+\Delta X)\cdot g(X)} \qquad \text{with } C = X\,.\end{aligned}$$

Now the general formula follows from the *Product Rule*.

Given two functions f, g differentiable at $X = A$, we have:

$$\boxed{\frac{d(f/g)}{dX}(A) = \frac{\frac{df}{dX}(A)\cdot g(A) - f(A)\cdot\frac{dg}{dX}(A)}{g(A)^2}}$$

provided $g(A) \neq 0$. So, ΔX goes to 0 and all Δs except for one are replaced with d's:

Theorem 4.7.7: Quotient Rule for Derivatives

For any two functions f, g differentiable at X, we have:

$$\boxed{\frac{d(f/g)}{dX} = \frac{f(x)\cdot\frac{dg}{dX} - \frac{df}{dX}\cdot g}{g^2}}$$

provided $g(X) \neq 0$.

Proof.

We represent the reciprocal of g as a composition:

$$z = \frac{1}{g(X)} \implies z = \frac{1}{y},\ y = g(X) \implies \frac{dz}{dy} = -\frac{1}{y^2},\ \frac{dy}{dX} = \nabla g(X) \implies \frac{dz}{dX} = -\frac{1}{g(X)^2}\nabla g(X)\,,$$

by the *Chain Rule*. Now the general formula follows from the *Product Rule*.

The formula is similar to the *Product Rule* in the sense that it also involves *cross-multiplication*:

	functions	derivatives
first	f	∇f
second	g	∇g

$$\longrightarrow \nabla\left(\frac{f}{g}\right) = \frac{f\nabla g - \nabla f g}{g^2}$$

Similar to the previous theorem, either term in the numerator is the product of a scalar function and a vector field. Their sum is a vector field and it's still a vector field when we divide by a scalar function.

This is the summary of the four properties re-stated in the gradient notation:

SR:	$\nabla(f+g) = \nabla f + \nabla g$	CMR:	$\nabla(kf) = k\,\nabla f$	for any real k
PR:	$\nabla(fg) = \nabla f\, g + f\,\nabla g$	QR:	$\nabla(f/g) = \dfrac{\nabla f\, g - f\,\nabla g}{g^2}$	wherever $g \neq 0$

4.8. *The gradient is perpendicular to the level curves*

The result we have been alluding to is that *the direction of the gradient is the direction of the fastest growth of the function*. It is proven later but here we just consider the relation between the gradient and the level curves, i.e., the curves of constant value, of the function.

We start with simple observations:

Theorem 4.8.1: Level Sets and Differences

Suppose a function $z = f(X)$ of several variables is defined at the adjacent nodes X and $X + \Delta X \neq X$ of a partition. Then, if these two nodes lie within a level set of $z = f(X)$, i.e., $f(X) = f(X + \Delta X)$, then

$$\Delta f(A) = 0\,,$$

where A is the secondary node of this edge.

Theorem 4.8.2: Level Sets and Difference Quotients

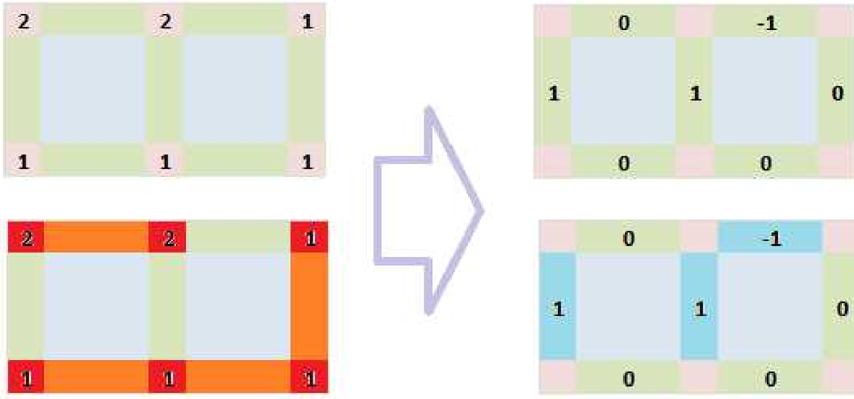
Suppose a function $z = f(X)$ of several variables is defined at the adjacent nodes X and $X + \Delta X \neq X$ of a partition. Then, if these two nodes lie within a level set of $z = f(X)$, i.e., $f(X) = f(X + \Delta X)$, then

$$\frac{\Delta f}{\Delta X}(A) = 0\,,$$

where A is the secondary node of this edge.

Example 4.8.3: spreadsheet

A function defined at the nodes on the plane is shown in the first column with its level curve visualized:



In the second column, the difference quotient is computed and then below it is visualized. This curve and this vector are perpendicular.

Exercise 4.8.4

Consider other possible arrangements of the values of the function and confirm the conjecture.

Example 4.8.5: parallel

In the familiar example of a plane:

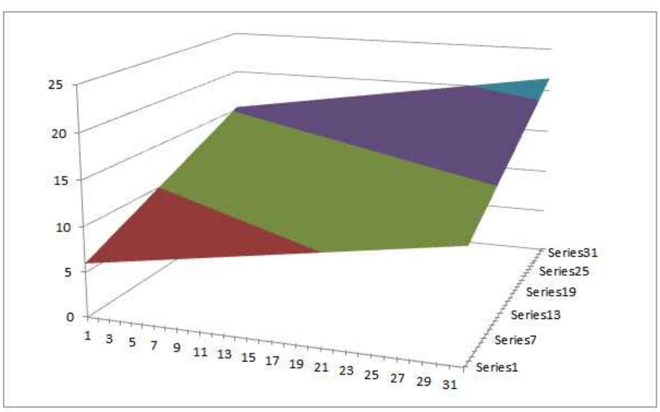
$$f(x,y) = 2x + 3y\,,$$

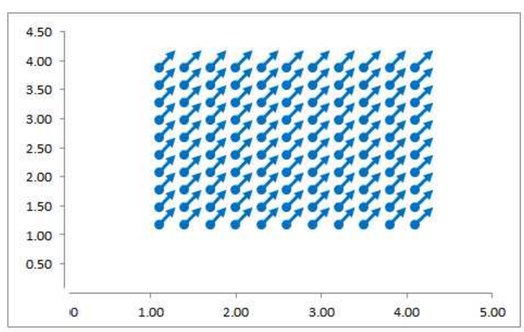
the gradient is a constant vector field:

$$\nabla f(x,y) =< 2, 3 > \ .$$

Meanwhile, its level curves are parallel straight lines:

$$2x + 3y = c\,.$$





The slope is $-2/3$ which makes them *perpendicular* to gradient vector $< 2, 3 >$!

We then conjecture that *the gradient and the level curves are perpendicular to each other.*

Let's consider a general *linear function* of two variables:

$$z = f(x,y) = c + m(x-a) + n(y-b)\,,$$

and $M = \nabla f =< m, n >$ the gradient of f. Let's pick a simple vector $D =< -n, m >$ perpendicular to $M =< m, n >$. Consider this straight line with D as a direction vector:

$$F(t) = (a,b)+ < -n, m > t\,.$$

We substitute it into f:

$$f(F(t)) = f(a - nt, b + mt) = c + m(-nt) + n(mt) = c\,.$$

The composition is constant and, therefore, the line stays within a level curve of f. The conjecture is confirmed.

Example 4.8.6: paraboloid

In the familiar example of a circular paraboloid:

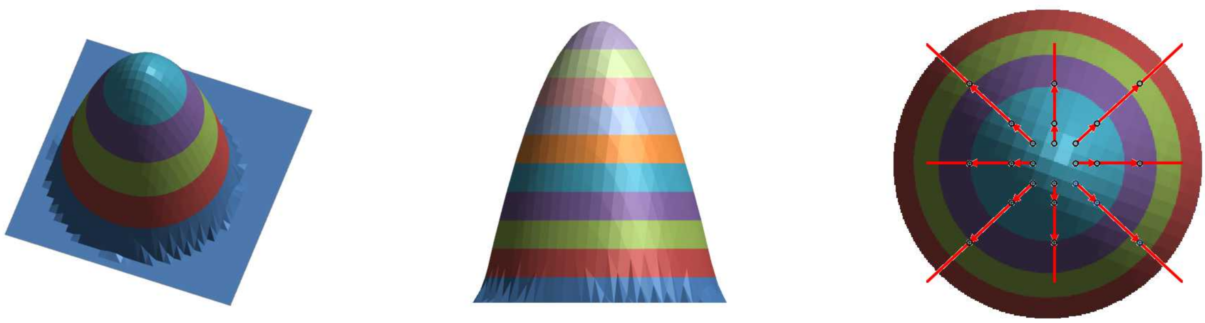
$$f(x, y) = x^2 + y^2\,,$$

the gradient consists of the radial vectors:

$$\nabla f(x, y) =< 2x, 2y > \ .$$

Meanwhile, its level curves are circles:

$$x^2 + y^2 = c \text{ for } c > 0\,.$$



The radii of a circle are known (and are seen above) to be *perpendicular* to the circle!

We need to make our conjecture precise before proving it. First, the level curves of a function of two variables aren't necessarily *curves*. They are just *sets* in the plane. For example, when f is constant, all the level sets are empty but one which is the whole plane:

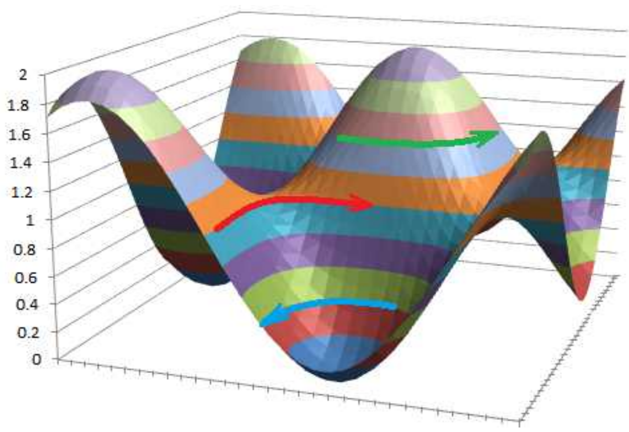
$$f(x, y) = c \implies \big\{(x, y) : f(x, y) = b\big\} = \emptyset \text{ when } b \neq c \ \text{ and } \ \big\{(x, y) : f(x, y) = c\big\} = \mathbf{R}^2\,.$$

Furthermore, even when a level curve is a curve, it's an implicit curve and isn't represented by a function.

Warning!

This doesn't mean that the path of F *is* a level set but simply its subset. The question of when exactly level curves are curves is addressed elsewhere.

How do we sort this out? Just as before, we study the terrain by taking these hikes – parametric curves – and this time we choose an easy one: no climbing. We stay at the *same elevation*:

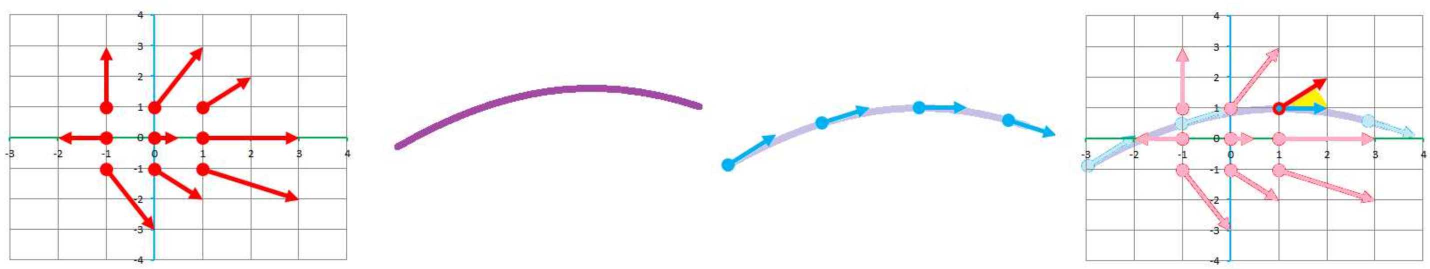


In other words, our function $z = f(X)$ does not change along this parametric curve $X = F(t)$, i.e., their composition is constant:

$$f(F(t)) = \text{ constant } .$$

We can go slow or fast and we can go in either direction.

Second, what do we mean when we say a parametric curve and a vector are perpendicular to each other? The direction of a curve at a point is its *tangent vector* at that point, by definition!



We are then concerned with:

$$\text{the angle between } \nabla f(A) \ \text{ and } \ F'(a), \text{ where } A = F(a)\,,$$

and, therefore, with their *dot product*:

$$\nabla f(F(a)) \cdot F'(a)\,.$$

Is it zero? But we just saw this expression in the last section! It's the right-hand side of the *Chain Rule*:

$$(f \circ F)'(a) = \nabla f(F(a)) \cdot F'(a)\,.$$

Why is the left-hand side zero? Because it's the derivative of a constant function! Indeed, the path of F lie within a level curve of f. So, we have:

$$0 = \frac{d}{dt} f(F(t))\bigg|_{t=a} = (f \circ F)'(a) = \nabla f(F(a)) \cdot F'(a)\,.$$

So, we have demonstrated that level curves and the gradient vectors are perpendicular:

$$\nabla f(A) \perp F'(a)\,.$$

What remains is just some caveats. First, the functions have to be differentiable for the derivatives to make sense. Second, neither of these derivatives should be zero or the angle between them will be undefined ($\nabla f(A) \neq 0$ and $F'(a) \neq 0$).

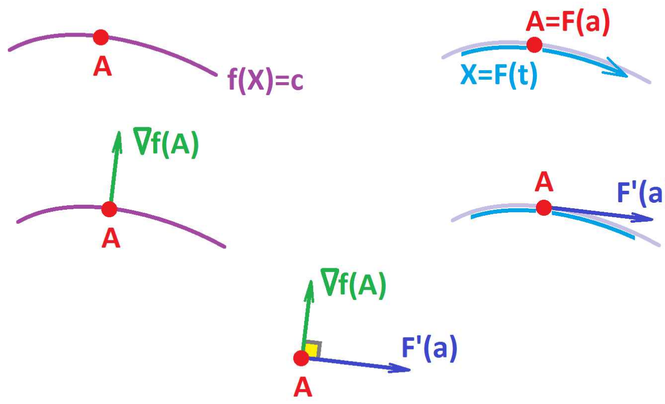
Theorem 4.8.7: Level Curves Are Perpendicular to Gradient

Suppose a function of several variables $z = f(X)$ is differentiable at $X = A$ and a parametric curve $X = F(t)$ is differentiable on an open interval I that contains a with $F(a) = A$. Then, if the path of $X = F(t)$ lies within a level set

of $z = f(X)$, then

$$\boxed{\frac{df}{dX}(A) \perp \frac{dF}{dt}(a)}$$

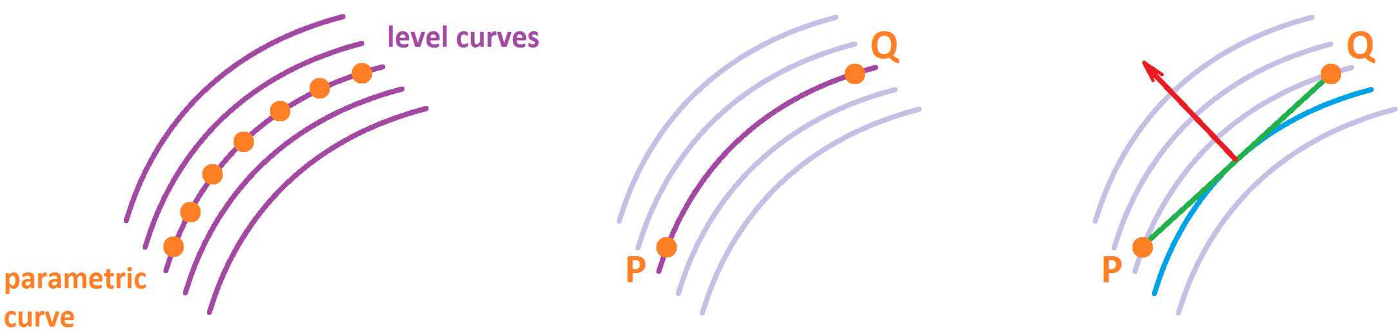
provided both are non-zero.



Exercise 4.8.8

What about the converse?

We now demonstrate the following result that mixes the discrete and the continuous:



Corollary 4.8.9: Level Curves Are Perpendicular to Gradient

Suppose a function of several variables $z = f(X)$ is differentiable on an open set U in $\mathbf{R}^n$. Suppose a parametric curve $X = F(t)$ is defined at adjacent nodes t and $t + \Delta t$ of a partition. Suppose the points $P = F(t)$ and $Q = F(t + \Delta t)$ are distinct and lie within a level set of $z = f(X)$, i.e., $f(P) = f(Q)$, and the segment PQ between them lies entirely within U. Then, for some point A on PQ and a secondary node a of $[t, t + \Delta t]$, we have:

$$\frac{df}{dX}(A) \perp \frac{\Delta F}{\Delta t}(a)\,,$$

provided the gradient is non-zero in U.

Proof.

Let $z = L(t)$ be the linear parametric curve with $L(t) = P$ and $L(t + \Delta t) = Q$. Then,

$$\frac{dL}{dt}(a) = \frac{\Delta F}{\Delta t}(a)\,,$$

for any choice of a secondary node a of the interval $[t, t + \Delta t]$. We define a new numerical function

defined at the nodes t and $t + \Delta t$:

$$h = f \circ F \,.$$

Then by the *Mean Value Theorem*, there is such a secondary node a that:

$$\frac{\Delta h}{\Delta t}(a) = \frac{dh}{dt}(a)\,.$$

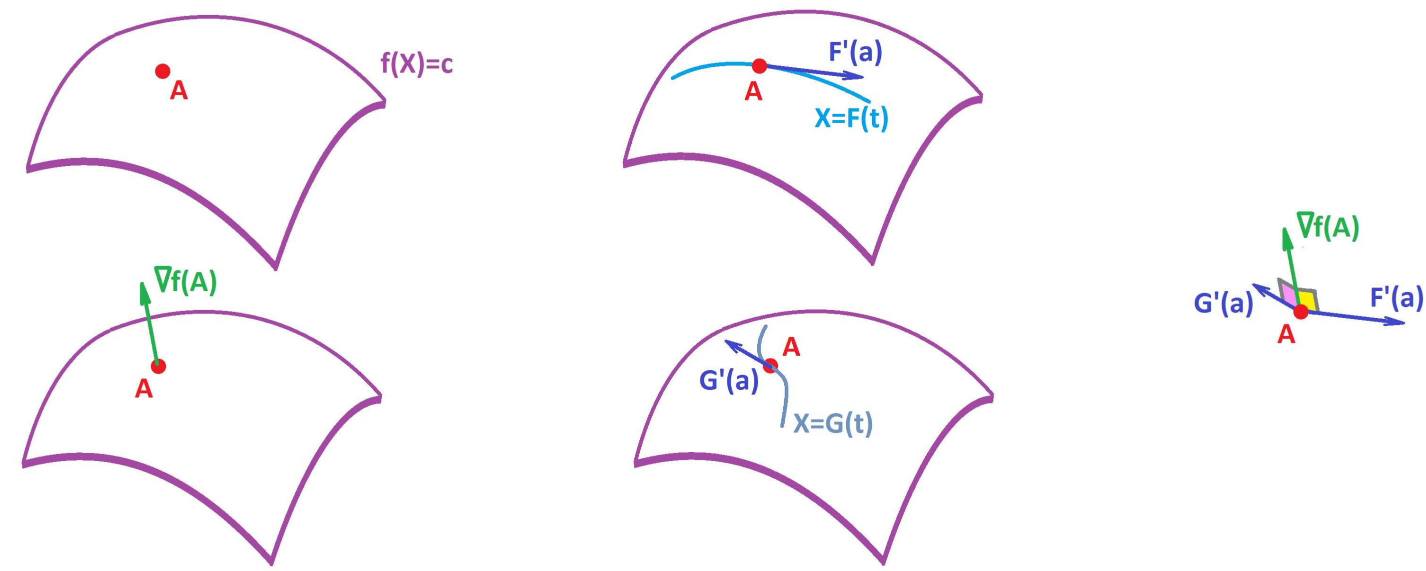
Since the former is zero by the assumption, we can apply the *Chain Rule* and conclude the following about the latter:

$$0 = \frac{dh}{dt}(a) = \frac{df}{dX}(A) \cdot \frac{dL}{dt}(a) = \frac{df}{dX}(A) \cdot \frac{\Delta F}{\Delta t}(a)\,,$$

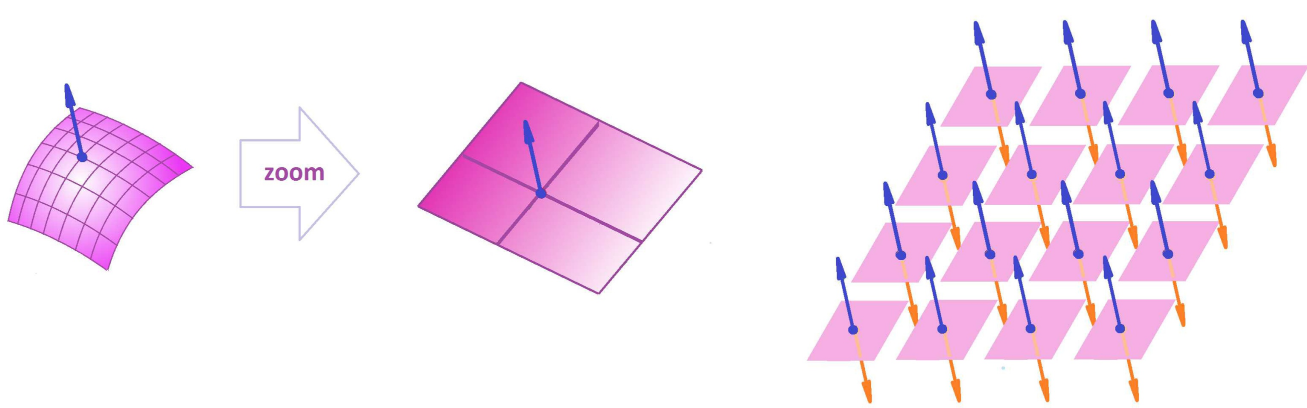
where $A = L(a)$.

The theorem remains valid no matter how a parametric curve traces the level curve as long as it doesn't stop. There are then only two main ways – back and forth – that a parametric *curve* can follow the level *curve*.

But wait a minute, the theorem doesn't speak exclusively of functions of *two* variables? It seems to apply to level *surfaces* of functions of three variables. Indeed. The basic idea is the same: A parametric curve perpendicular to the gradient, even though there are infinitely many directions for the curve to go through the point.



With all the variety of angles between their tangents, they all have the same angle with the gradient. In this exact sense we speak of the gradient being perpendicular to the level surface.



This result is a free gift courtesy of abstract thinking and the vector notation!

Example 4.8.10: radial vector field

The level surfaces of the radial vector field,

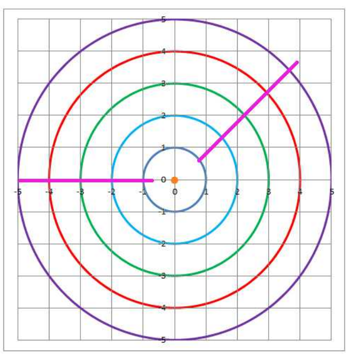
$$V(x,y,z) = < x,y,z > ,$$

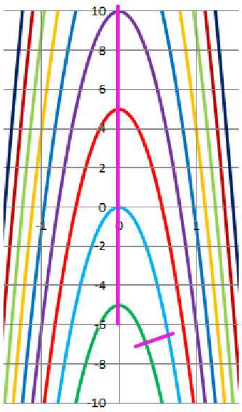
as well as the vector field of the gravitation,

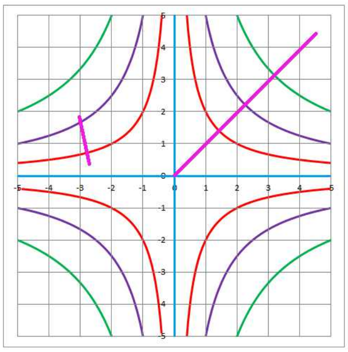
$$W(x,y,z) = -\frac{c}{|| < x,y,z > ||^3} < x,y,z > ,$$

are concentric spheres. The gradient vectors point away from the origin in the former case and towards it in the latter. One can imagine how, no matter what path you take on the surface of the Earth, your body will point away from the center.

With this theorem we can interpret the idea that *the gradient points in the direction of the fastest growth of the function.* Indeed, it suggests the shortest path toward the "next" level curve:





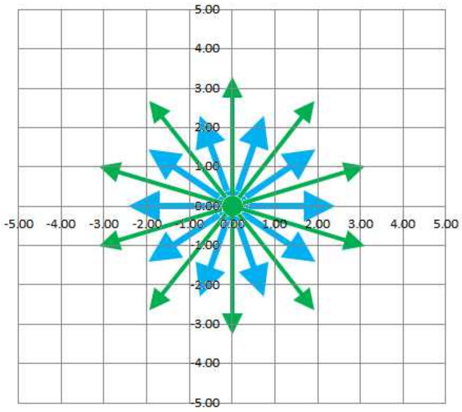


This informal explanation isn't good enough anymore. We will make the terms in this statement fully precise next.

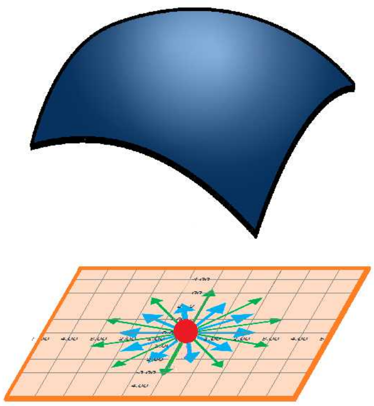
4.9. Monotonicity of functions of several variables

Suppose $z = f(X)$ is a function of n variables. Suppose A is a point in $\mathbf{R}^n$. Then $V = \nabla f(A) \neq 0$ is a vector. As such it has a *direction*. More precisely, the direction of V is its normalization, the unit vector $V/||V||$. Thus, the first part of the statement "the gradient points in the direction of the fastest growth of the function" is well understood. But what does "the direction of the fastest growth of the function" mean?

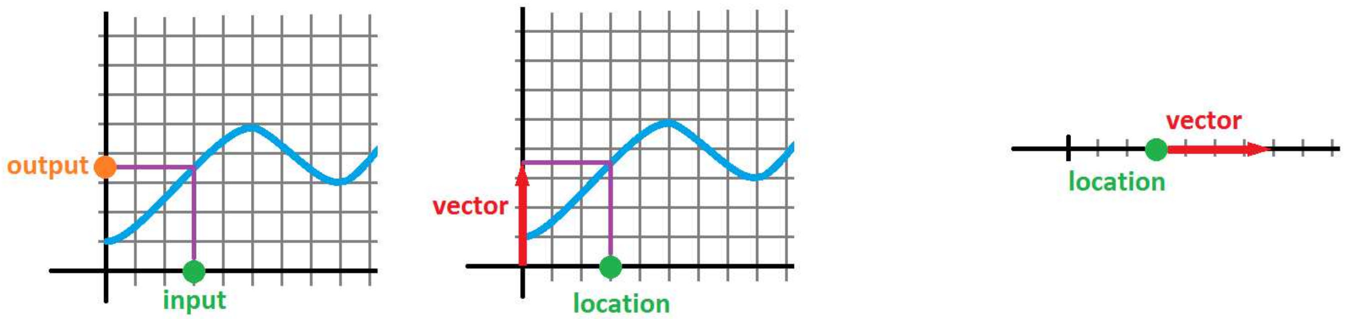
First, the gradient will be chosen from *all possible directions*, i.e., all unit vectors:



Then, what does the "growth of the function in the direction of a unit vector" mean?



Let's first take a look at dimension $n = 1$. There are only *two* unit vectors, i and $-i$, along the x-axis. Therefore, if $f'(A) > 0$, then i is the direction of the fastest growth; meanwhile, if $f'(A) < 0$, it's $-i$.



For higher dimensions, we certainly know what this statement means when the direction coincides with the direction of one of the axes: it's the partial derivative (vectors i, $-i$, j, $-j$ etc.). However, if we are exploring the terrain represented by a function of two variables, going only north-south or east-west is not enough. The idea comes from the earlier part of this section: we, again, take various trips around this terrain. This time we don't have to go far or follow any complex routs: we'll go along straight lines. Also, in order to compare the results, we will travel at the same speed, 1, during all trips.

We will consider all parametric curves $X = F_U(t)$ that

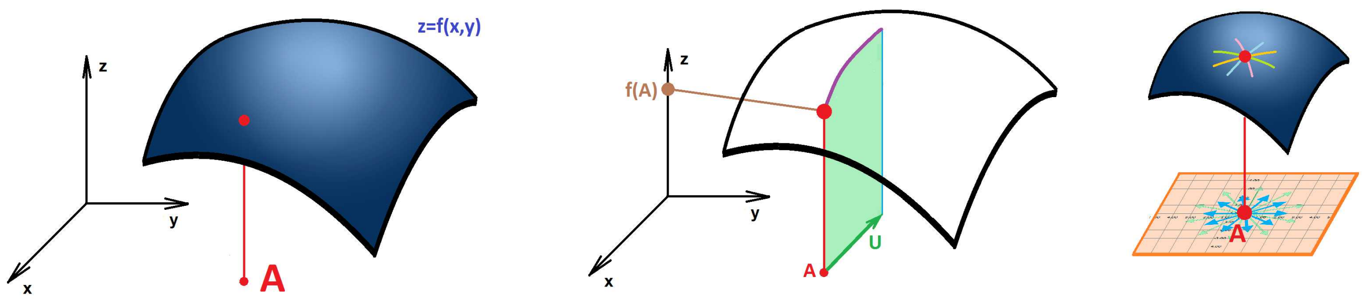
- start at $X = A$, i.e., $F_U(0) = A$,
- are linear, i.e., $F_U(t) = A + tU$, and, furthermore,
- have unit direction vector $||U|| = 1$.

Warning!

It is safe to disregard non-linear parametric curves only under the assumption that f is differentiable.

Now we compare the rate of growth of f along these parametric curves by considering their composition with f:

$$h_U(t) = f(F_U(t)) .$$



So, the rate of growth we are after is this:

$$h'_U(0) = \frac{d}{dt} f(F_U(t))\Big|_{t=0} = \nabla f(F_U(t)) \cdot F'_U(t)\Big|_{t=0} = \nabla f(A) \cdot U\,,$$

according to the *Chain Rule*. There is a convenient term for this quantity.

Definition 4.9.1: directional derivative

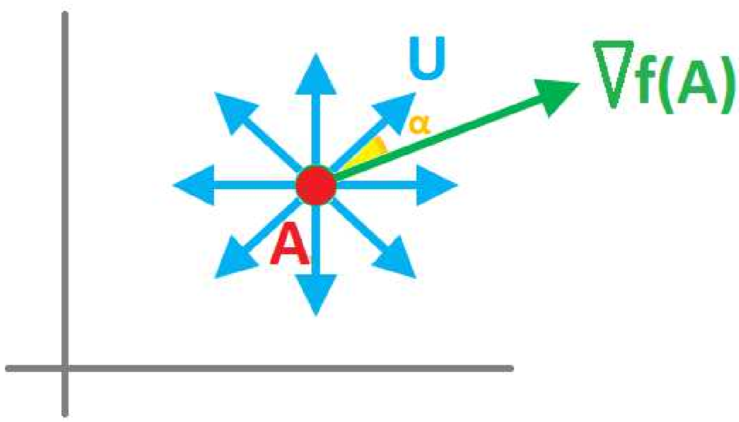
The *directional derivative of a function* $z = f(X)$ *at point* $X = A$ *in the direction of a unit vector* U is defined to be

$$\boxed{D_U(f, A) = \nabla f(A) \cdot U}$$

We continue the above computation:

$$D_U(f, A) = ||\nabla f(A)|| \cdot ||U|| \cos\alpha = ||\nabla f(A)|| \cos\alpha\,,$$

where α is the angle between $\nabla f(A)$ and U. As the gradient is known and fixed, the directional derivative in a particular direction depends on its angle with the gradient, as expected:



Now, this expression is easy to maximize over the vectors U. What direction, i.e., a unit vector U, provide the highest value of $D_U(f, A)$? Only $\cos\alpha$ matters! And it reaches is maximum values, which is 1, at $\alpha = 0$. In other words, the maximum is reached when the direction coincides with the gradient!

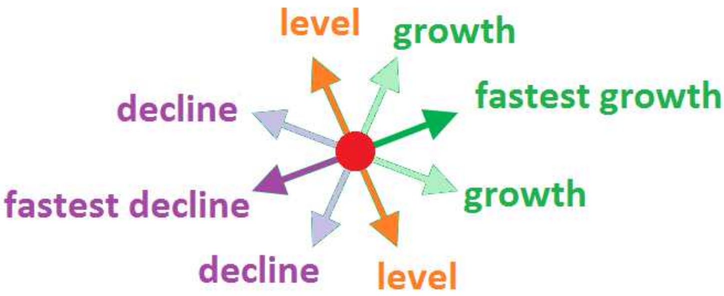
Theorem 4.9.2: Monotonicity Theorem

Suppose $z = f(X)$ *is a function of* n *variables differentiable at a point* A *in* $\mathbf{R}^n$*. Then the directional derivative* $D_U(f, A)$ *reaches it maximum in the direction* U *of the gradient* $\nabla f(A)$ *of* f *at* A*; this maximum value is* $||\nabla f(A)||$.

Exercise 4.9.3

Show that the theorem for dimension 1 reveals increasing and decreasing behavior.

This is the summary of the theorem and the rest of the analysis:



Exercise 4.9.4

Explain the diagram.

An alternative definition is similar to the one for the derivative of a numerical function:

Theorem 4.9.5: Directional Derivative

The directional derivative of a function $z = f(X)$ at point $X = A$ in the direction of a unit vector U is also found as the following limit:

$$\boxed{D_U(f, A) = \lim_{h\to 0} \frac{f(A + hU) - f(A)}{h}}$$

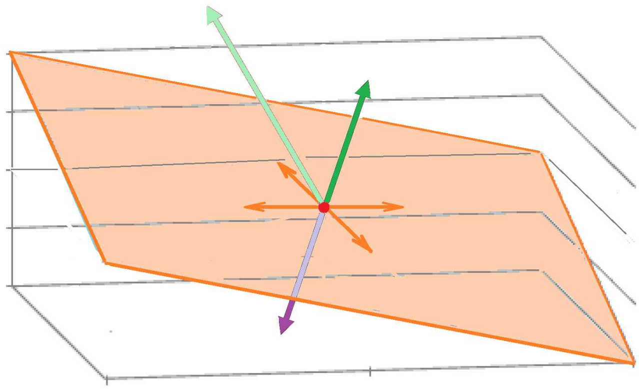
Exercise 4.9.6

Prove the theorem.

Exercise 4.9.7

Represent each partial derivative as a directional derivative.

This is what happens with functions of 3 variables:



All vectors on one side of the level surface are the directions of increasing values of the function and all on the other side are the directions of decreasing values.

4.10. *Differentiation and anti-differentiation*

Let's review the algebraic properties of differentiation of functions of several variables. The properties are the same as before!

Theorem 4.10.1: Algebra of Gradients

For any differentiable functions, we have in the gradient notation:

SR:	$\nabla(f+g)=\nabla f+\nabla g$	*CMR:*	$\nabla(cf)=c\nabla f$	*real* c
PR:	$\nabla(fg)=\nabla f\, g+f\nabla g$	*QR:*	$\nabla(f/g)=\dfrac{\nabla f\, g-f\nabla g}{g^2}$	$g\neq 0$
CR1:	$(f\circ F)'=\nabla f\cdot F'$	*CR2:*	$(g\circ f)'=g'\nabla f$	

The *Mean Value Theorem* (Chapter 2DC-5) will help us to derive facts about the function from the facts about its gradient. For example:

info about f		info about ∇f
f is constant	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	∇f is zero
f is linear	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	∇f is constant

Are these arrows reversible? If the derivative of the function is zero, does it mean that the function is constant? At this time, we have a tool to prove this fact.

Consider this simple statement about terrains:

- "If there is no sloping anywhere in the terrain, it's flat".

If a function $y=f(x)$ represents the position, we can restate this mathematically:

Theorem 4.10.2: Constant Rule For Differences

If a function defined at the nodes of a partition of a cell in $\mathbf{R}^n$ *has a zero difference throughout the partition, then this function is constant over the nodes; i.e.,*

$$\boxed{\Delta f\,(C)=0\ \Longrightarrow\ f=\ \textit{constant}}$$

Proof.

If X and Y are two nodes connected by an edge with a secondary node C, then we have:

$$\Delta f\,(C)=0\ \Longrightarrow\ f(X)-f(Y)=0\ \Longrightarrow\ f(X)=f(Y)\,.$$

In a cell, any two nodes can be connected by a sequence of adjacent nodes, with no change in the value of f.

Theorem 4.10.3: Constant Rule For Difference Quotients

If a function defined at the nodes of a partition of a cell in $\mathbf{R}^n$ *has a zero difference quotient throughout the partition, then this function is constant over the nodes; i.e.,*

$$\boxed{\frac{\Delta f}{\Delta X}(C)=0\ \Longrightarrow\ f=\ \textit{constant}}$$

Theorem 4.10.4: Constant Rule For Derivatives

If a function differentiable on an open path-connected set I in $\mathbf{R}^n$ has a zero gradient for all X in I, then this function is constant on I; i.e.,

$$\boxed{\frac{df}{dX} = 0 \implies f = \text{ constant}}$$

Proof.

Suppose two points A, B inside I are given. Then there is a differentiable parametric curve $X = P(X)$ with its path that goes from A to B and lies entirely in I:

$$P(a) = A,\ P(b) = B,\ P(t) \text{ in } I\,.$$

Define a new *numerical function*:

$$h(t) = f(P(t))\,.$$

Then, by the *Chain Rule* we have:

$$\frac{dh}{dt}(t) = \frac{d}{dt}\big(f(P(t))\big) = \nabla f\,(P(t)) \cdot F'(t) = 0 \cdot F'(t) = 0\,.$$

Then, by the corollary to the *Mean Value Theorem* in Volume 2 (Chapter 2DC-5), h is a constant function. In particular, we have

$$f(A) = f(B)\,.$$

We will see later that the differentiability requirement is unnecessary.

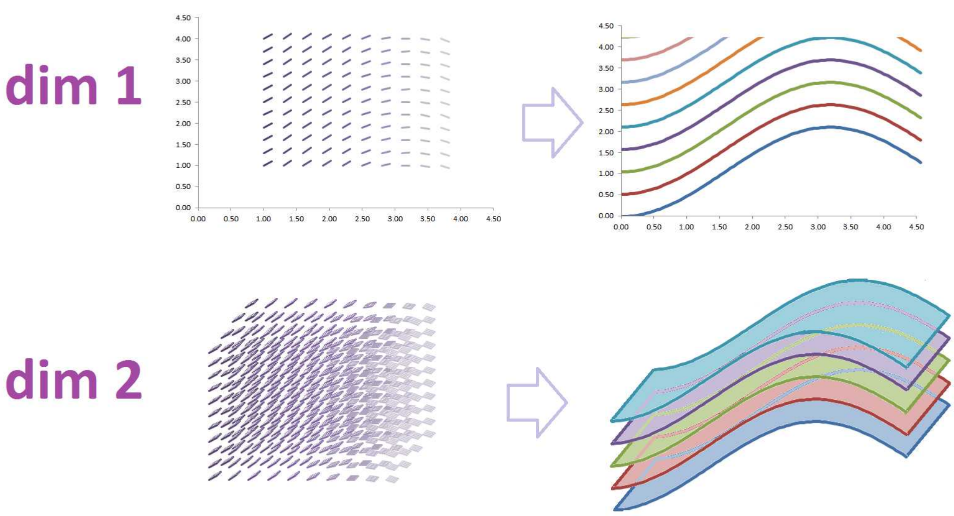
Exercise 4.10.5

What if $\nabla f = 0$ on a set that isn't path-connected? Is it still true that

$$\nabla f = 0 \implies f = \text{ constant ?}$$

Just as in dimension 1, the openness of the domain is crucial.

The problem then becomes one of recovering the function f from its derivative (i.e., gradient) ∇f, the process we have called *anti-differentiation*. In other words, we reconstruct the function from a "field of tangent lines or planes":



Now, even if we can recover the function f from it derivative ∇f, there many others with the same derivative, such as $g = f + C$ for any constant vector C. Are there others? No.

Theorem 4.10.6: Anti-differentiation For Differences

If two functions defined at the nodes of a partition of a cell in $\mathbf{R}^n$ have the same difference, they differ by a constant; i.e.,

$$\boxed{\Delta f\,(C) = \Delta g\,(C) \implies f(X) - g(X) = \text{ constant}}$$

Theorem 4.10.7: Anti-differentiation For Difference Quotients

If two functions defined at the nodes of a partition of a cell in $\mathbf{R}^n$ have the same difference quotient, they differ by a constant; i.e.,

$$\boxed{\frac{\Delta f}{\Delta X}\,(C) = \frac{\Delta g}{\Delta X}\,(C) \implies f(X) - g(X) = \text{ constant}}$$

Theorem 4.10.8: Anti-differentiation For Derivatives

If two functions differentiable on an open path-connected set I in $\mathbf{R}^n$ have the same gradient, they differ by a constant; i.e.,

$$\boxed{\frac{df}{dX} = \frac{dg}{dX} \implies f - g = \text{ constant}}$$

Proof.

Define
$$h(X) = f(X) - g(X)\,.$$
Then, by SR, we have:
$$\nabla h\,(X) = \nabla\,(f(X) - g(X)) = \nabla f\,(X) - \nabla g\,(X) = 0\,,$$
for all X. Then h is constant, by the *Constant Theorem*.

Geometrically,
$$\nabla f = \nabla g \implies f - g = \text{ constant }\,,$$
means that the graph of f shifted vertically gives us the graph of g.

We can cut the list of algebraic rules down to the most important ones:

Linearity Rule:	$\nabla(\lambda f + \mu g) = \lambda \nabla f + \mu \nabla g$ for all real λ, μ
Chain Rule:	$(f \circ F)' = \nabla f \cdot F'$

4.11. *When is anti-differentiation possible?*

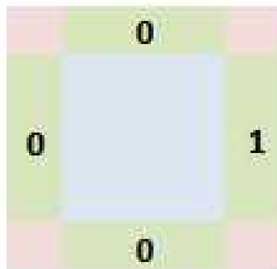
Back to the discrete case. We concentrate on dimension 2 here.

Recall the diagram of partial differentiation of a function of two variables:

$$\begin{array}{ccc} & f & \\ \swarrow_x & & {}_y\searrow \\ \Delta f =<\ \Delta_x f & , & \Delta_y f\ > \end{array}$$

It produces the difference and the difference quotient both of which are functions defined at the secondary nodes of a partition.

In contrast to dimension 1, it is easy to think of a 1-form that isn't the difference of any function:



Exercise 4.11.1

Prove that there is no such function.

Definition 4.11.2: exact function

A function G defined on the secondary nodes of a partition is called *exact* if $\Delta f = G$, for some function f defined on the nodes of the partition. When the secondary nodes aren't specified, we speak of an exact 1-form.

In fact, choosing numbers at random is likely to produce non-exact form.

Definition 4.11.3: gradient

A vector field F defined on the secondary nodes of a partition is called *gradient* if $F(N) \cdot N = G(N)$ for some exact function G and any secondary node N.

Not all vector fields are gradient; the example we saw was $F(x, y) =< y, -x >$.

Anti-differentiation is reversing the arrows in the above diagram:

$$\begin{array}{ccc} & f =? & \\ \nearrow_x & & {}_y\nwarrow \\ <\ p & , & q\ > \end{array}$$

How do we know that a given function defined at the secondary nodes is exact? In other words, is it the difference of some function? We have previously solved this problem by finding such a function. The examples required producing the recursive formulas for x and y and then matching their applications in reverse order. The methods only work when the functions are simple enough.

Can we know in advance?

The familiar theorem below gives us a better tool. Surprisingly, this tool is *further partial differentiation*.

We continue the above diagram:

$$\begin{tikzcd}
 & & f \arrow[dl, "x"'] \arrow[dr, "y"] & & \\
\Delta f =< & \Delta_x f \arrow[dl, "x"'] \arrow[dr, "y"] & , & \Delta_y f \arrow[dl, "x"'] \arrow[dr, "y"] & > \\
\Delta^2_{xx} f & & \color{red}{\Delta^2_{yx} f = \Delta^2_{xy} f} & & \Delta^2_{yy} f
\end{tikzcd}$$

Recall the following result from Chapter 3:

> **Theorem 4.11.4: Discrete Clairaut's Theorem**
>
> *Over a partition in $\mathbf{R}^n$, first, the mixed second differences with respect to any two variables are equal to each other:*
>
> $$\boxed{\Delta^2_{yx} f = \Delta^2_{xy} f}$$
>
> *and, second, the mixed second difference quotients are equal to each other:*
>
> $$\boxed{\frac{\Delta^2 f}{\Delta y \Delta x} = \frac{\Delta^2 f}{\Delta x \Delta y}}$$

Thanks to this theorem we can draw conclusions from the assumption that we face a difference. So, the plan is, instead of trying to reverse the arrows in the first row of the diagram, we continue down and see whether we have a match:

$$\Delta_y p = \Delta_x q\,.$$

As a summary, consider an arbitrary function on secondary nodes. It has arbitrary component functions, with no relations between them whatsoever! Everything changes once we make the assumption that it is exact. The theorem ensures that the diagram of the differences of the component functions of some $V =< p, q >$, on left, turns – under this assumption – into something *rigid*, on right:

$$\begin{tikzcd}
 & \cdot\cdot & & & & f \arrow[dl, "x"'] \arrow[dr, "y"] & \\
V =< p \arrow[dr, "y"] & , & q > \arrow[dl, "x"'] & \to & p = \Delta_x f \arrow[dr, "y"] & & q = \Delta_y f \arrow[dl, "x"'] \\
 & \Delta_y p \ldots \Delta_x q & & & & \Delta_y p = \Delta^2_{yx} f = \Delta^2_{xy} f = \Delta_x q &
\end{tikzcd}$$

This rigidity of the diagram means that the two trips from the top to the bottom produce the same result. We have described this property as *commutativity*. Indeed, it's about interchanging the order of the two operations of partial differentiation:

$$\Delta_x \Delta_y = \Delta_y \Delta_x\,.$$

> **Theorem 4.11.5: Exactness Test For Dimension 2**
>
> *If G is exact on a rectangle on the xy-plane with component functions p and q, then*
>
> $$\boxed{\Delta_y p = \Delta_x q}$$

Corollary 4.11.6: Gradient Test For Dimension 2

Suppose a vector field V is defined on the secondary nodes of a partition of a rectangle in the xy-plane with component functions p and q. If V is gradient, then

$$\boxed{\frac{\Delta p}{\Delta y} = \frac{\Delta q}{\Delta x}}$$

The quantity that vanishes when the function is gradient is called its *rotor*. It is a real-valued function defined on the faces of the partition,

$$\Delta_y p - \Delta_x q\,.$$

For *three variables*, we just consider two at a time with the third kept fixed.

Below is the diagram of the partial differences for three variables with only the mixed partial differences shown:

$$\begin{array}{ccccccccccc}
 & & \longleftarrow & \longleftarrow & \longleftarrow & f & \longrightarrow & \longrightarrow & \longrightarrow & & \\
 & \swarrow_x & & & & \downarrow_y & & & & {}_z\searrow & \\
\Delta_x f & & & & & \Delta_y f & & & & & \Delta_z f \\
\downarrow_y & {}_z\searrow & & & \swarrow_x & & {}_z\searrow & & & \swarrow_x & \downarrow_y \\
\color{red}{\Delta^2_{yx} f} & & \color{green}{\Delta^2_{zx} f} & \color{red}{\Delta^2_{xy} f} & & & & \color{blue}{\Delta^2_{zy} f} & \color{green}{\Delta^2_{xz} f} & & \color{blue}{\Delta^2_{yz} f}
\end{array}$$

The six that are left are paired up according to *Discrete Clairaut's Theorem* above.

Let's assume:

$$\begin{aligned} V &=< p, q, r > \\ &=< \Delta_x f, \Delta_y f, \Delta_z f >= \Delta f \end{aligned}$$

Then we can trace the differences of the component functions in the diagram:

$$\begin{array}{ccccccccc}
p & & & & q & & & & r \\
\downarrow_y & {}_z\searrow & & \swarrow_x & & {}_z\searrow & & \swarrow_x & \downarrow_y \\
\color{red}{\Delta_y p} & & \color{green}{\Delta_z p} \;\; \color{red}{\Delta_x q} & & & & \color{blue}{\Delta_z q} \;\; \color{green}{\Delta_x r} & & \color{blue}{\Delta_y r}
\end{array}$$

We write down the results below.

Theorem 4.11.7: Exactness Test For Dimension 3

If G is exact on a partition of a box in the xyz-space with component functions p, q, and r, then

$$\boxed{\Delta_y p = \Delta_x q,\ \ \Delta_z q = \Delta_y r,\ \ \Delta_x r = \Delta_z p}$$

Corollary 4.11.8: Gradient Test For Dimension 3

Suppose a vector field V is defined on the secondary nodes of a partition of a box in the xyz-space with component functions p, q, and r. If V is gradient, then

$$\boxed{\frac{\Delta p}{\Delta y} = \frac{\Delta q}{\Delta x}\,,\ \frac{\Delta q}{\Delta z} = \frac{\Delta r}{\Delta y}\,,\ \frac{\Delta r}{\Delta x} = \frac{\Delta p}{\Delta z}}$$

There is a simple pattern in these formulas. First the variables and the components are arranged around a triangle:

$$\begin{array}{ccccc} & & x & & \\ & \nearrow & & \searrow & \\ z & & \leftarrow & & y \end{array} \qquad \begin{array}{ccccc} & & p & & \\ & \nearrow & & \searrow & \\ r & & \leftarrow & & q \end{array}$$

Then one of the variables is omitted and the difference over the other two is set to 0:

$$\begin{array}{ccc} & \cdot & \\ \bullet & \Delta_z q = \Delta_y r & \bullet \end{array} \qquad \begin{array}{ccc} \bullet & & \\ & \Delta_y p = \Delta_x q & \\ \cdot & & \bullet \end{array} \qquad \begin{array}{ccc} & & \bullet \\ & \Delta_x r = \Delta_z p & \\ \bullet & & \cdot \end{array}$$

The conditions of these two theorems, and their analogs in higher dimensions, put severe limitations on what functions can be exact.

4.12. When is a vector field a gradient?

We proceed to the continuous case.

Recall the diagram of partial differentiation of a function of two variables that produces the vector field of the gradient:

$$\begin{array}{ccccccc} & & & f & & & \\ & & \swarrow_x & & {}_y\searrow & & \\ \nabla f =< & f_x & & , & & f_y & > \end{array}$$

The following is the continuous analog of an exact 1-form:

Definition 4.12.1: gradient vector field

A vector field that is the gradient of some function of several variables is called a *gradient vector field*. This function is then called a *potential function* of the vector field.

Note that finding for a given vector field V a function f such that $\nabla f = V$ amounts to anti-differentiation as we try to reverse the arrows in the above diagram. The *Anti-differentiation Theorem* is an analog of several familiar result from Volume 2:

$$\nabla f = \nabla g \implies f - g = \text{constant}\,.$$

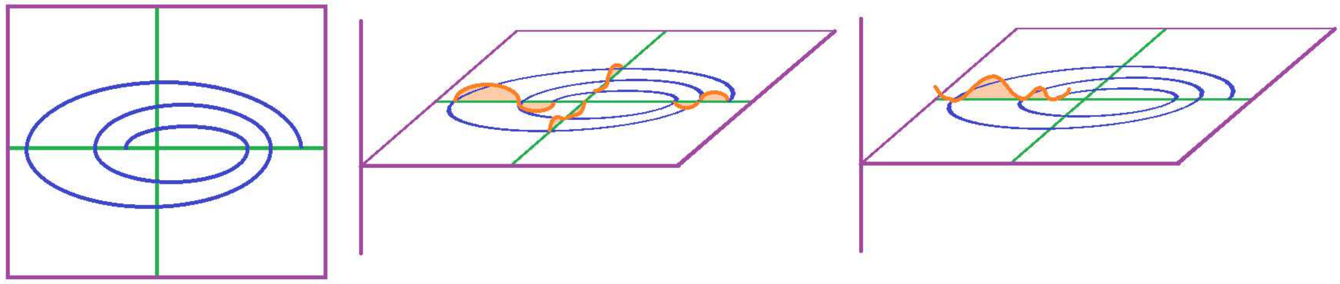
Corollary 4.12.2: Two Potential Functions

Any two potential functions of the same vector field defined on an open path-connected set differ by a constant within this set.

Not all vector fields are gradient. The example we saw was $V(x, y) =< y, -x >$. There are many more...

Example 4.12.3: spiral

Consider the spiral below. Can it be a level curve of a function of two variables?



Anti-differentiation is reversing the arrows in the above diagram:

$$\begin{array}{ccccc} & & f=? & & \\ & \nearrow_x & & {}_y\nwarrow & \\ < \ p & & , & & q \ > \end{array}$$

So, how do we know that a given vector field is gradient? We have previously solved this problem by finding, and trying and failing to find, a potential function for the vector field in dimension 2. The examples required integration with respect to both variables and then matching the results. The methods only work when the functions are simple enough.

Can we know in advance?

The familiar theorem below gives us such a tool. Just as in the last section with the discrete case, the answer is further *partial differentiation*. We continue the differentiation diagram:

$$\begin{array}{ccccccccc} & & & & f & & & & \\ & & & \swarrow_x & & {}_y\searrow & & & \\ & \nabla f =< & f_x & & , & & f_y & > & \\ & \swarrow_x & & {}_y\searrow & & \swarrow_x & & {}_y\searrow & \\ f_{xx} & & & & {\color{red}f_{yx}} = f_{xy} & & & & f_{yy} \end{array}$$

Warning!

The last row is the derivative of the vector field (of the gradient) somehow...

Recall the following result from Chapter 3 that gives us the equality of the mixed second derivatives:

Theorem 4.12.4: Clairaut's Theorem

The mixed second derivatives of a function f of two variables with continuous second partial derivatives at a point (a,b) are equal to each other; i.e.,

$$\boxed{f_{xy}(a,b) = f_{yx}(a,b)}$$

Thanks to this theorem we can draw conclusions from the assumption that a given vector field $V=<p,q>$ is gradient – as long as its component functions p and q are twice continuously differentiable. So, the plan is, instead of trying to reverse the arrows in the first row of the diagram and find f with $\nabla f = V$, we continue down and see whether we have a match:

$$p_y = q_x\,.$$

Example 4.12.5: testing

It's easy. For $V =< x, y >$, we have:

$$\left.\begin{array}{lll} p = x & \Longrightarrow & p_y = 0 \\ q = y & \Longrightarrow & q_x = 0 \end{array}\right\} \Longrightarrow \text{ match!}$$

The test is passed! So what? What do we conclude from that? Nothing.

On the other hand, $V =< y, -x >$, we have:

$$\left.\begin{array}{lll} p = y & \Longrightarrow & p_y = 1 \\ q = x & \Longrightarrow & q_x = -1 \end{array}\right\} \Longrightarrow \text{ no match!}$$

The test is failed. It's not gradient!

We draw no conclusion when the test is passed and when it isn't, we would still have to integrate to find out if it is gradient and, at the same time, try to find the gradient. Meanwhile the failure to satisfy the test proves the vector field is *not* gradient.

As a summary, consider an arbitrary vector field. It has arbitrary component functions, with no relations between them whatsoever! Everything changes once we make the assumption that this is a gradient vector field. The theorem ensures that the diagram of the derivatives of the component functions of an arbitrary vector field $V =< p, q >$ on left turns – under the assumption that it's gradient – into something *rigid* on right:

$$\begin{tikzcd} & & .. & & \\ & .. & & .. & \\ V =< p \arrow[dr, "y"'] & & , & & q > \arrow[dl, "x"] \\ & & p_y \, ... \, q_x & & \end{tikzcd} \quad \text{leading to} \quad \begin{tikzcd} & f \arrow[dl, "x"'] \arrow[dr, "y"] & \\ p = f_x \arrow[dr, "y"'] & & q = f_y \arrow[dl, "x"] \\ & p_y = f_{yx} = f_{xy} = q_x & \end{tikzcd}$$

This rigidity of the diagram means that the two trips from the top to the bottom produce the same result. We have described this property as *commutativity*. Indeed, it's about interchanging the order of the two operations of partial differentiation:

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x}\,.$$

Theorem 4.12.6: Gradient Test For Dimension 2

Suppose $V =< p, q >$ is a vector field with continuously differentiable on an open disk in $\mathbf{R}^2$ component functions p and q. If V is gradient, then

$$\boxed{p_y = q_x}$$

The quantity that vanishes when the vector field is gradient is called the *rotor* of the vector field. It is a function of two variables:

$$p_y - q_x\,.$$

We will see later how the rotor is used to measure how close the vector field is to being gradient.

For *three variables*, we just consider two at a time with the third kept fixed.

Below is the diagram of the partial derivatives for three variables with only the mixed derivatives shown:

$$\begin{array}{ccccccc}
\longleftarrow & \longleftarrow & \longleftarrow & f & \longrightarrow & \longrightarrow & \longrightarrow \\
\swarrow_x & & & \downarrow_y & & & {}_z\searrow \\
f_x & & & f_y & & & f_z \\
\downarrow_y \quad {}_z\searrow & & \swarrow_x & & {}_z\searrow & & \swarrow_x \quad \downarrow_y \\
f_{yx} \qquad f_{zx} & f_{xy} & & & & f_{zy} & f_{xz} \qquad f_{yz}
\end{array}$$

The six that are left are paired up according to Clairaut's theorem. If we choose

$$V =< p, q, r >=< f_x, f_y, f_z >= \nabla f \,,$$

we can trace the derivatives of the component functions in the diagram:

$$\begin{array}{ccccc}
p & & q & & r \\
\downarrow_y \quad {}_z\searrow & \swarrow_x & & {}_z\searrow \quad \swarrow_x & \downarrow_y \\
p_y \qquad p_z & q_x & & q_z \quad r_x & r_y
\end{array}$$

We write down the results below:

Theorem 4.12.7: Gradient Test For Dimension 3

Suppose $V =< p, q, r >$ is a vector field with continuously differentiable on an open ball in $\mathbf{R}^3$ component functions p, q, and r. If V is gradient, then

$$\boxed{p_y = q_x, \ q_z = r_y, \ r_x = p_z}$$

We have the same simple pattern. The variables and the components are arranged around a triangle:

$$\boxed{\begin{array}{ccc} & x & \\ \nearrow & & \searrow \\ z & \leftarrow & y \end{array}} \qquad \boxed{\begin{array}{ccc} & p & \\ \nearrow & & \searrow \\ r & \leftarrow & q \end{array}}$$

Then one of the variables is omitted and the rotor over the other two is set to 0:

$$\boxed{\begin{array}{ccc} & \cdot & \\ & & \\ \bullet & q_z = r_y & \bullet \end{array}} \qquad \boxed{\begin{array}{ccc} & \bullet & \\ & p_y = q_x & \\ \cdot & & \bullet \end{array}} \qquad \boxed{\begin{array}{ccc} & \bullet & \\ & r_x = p_z & \\ \bullet & & \cdot \end{array}}$$

All three of these quantities: $p_y - q_x, \ q_z - r_y, \ r_x - p_z$, vanish when the vector field is gradient. In order to have only one, we will use them as components to form a new vector field, called the *curl* of the vector field.

The conditions of these two theorems, and their analogs in higher dimensions, put severe limitations on what vector fields can be gradient. The source of these limitations is the topology of the Euclidean spaces of dimension 2 and higher. They are to be discussed later.

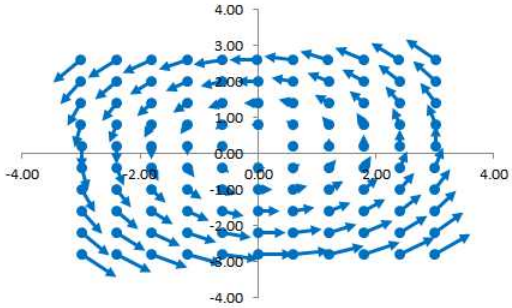
Back to dimension 2.

Example 4.12.8: rotation vector field

The rotation vector field,

$$V =< y, -x > \,,$$

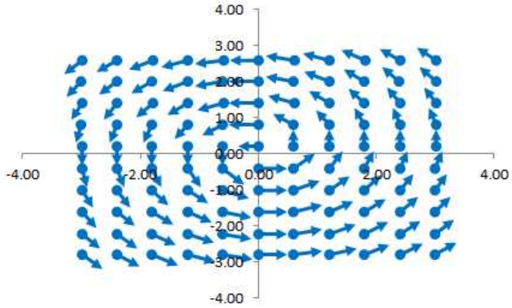
is not gradient as it fails the *Gradient Test*.



Let's consider its normalization:

$$U=\frac{V}{||V||}=\frac{1}{\sqrt{x^2+y^2}}<y\,,-x>=\left\langle\frac{y}{\sqrt{x^2+y^2}},\ -\frac{x}{\sqrt{x^2+y^2}}\right\rangle=<p,q>\ .$$

All vectors are unit vectors with the same directions as the last vector field:



Let's test the condition of the *Gradient Test*:

$$\begin{aligned}p_y&=\frac{\partial}{\partial y}\frac{y}{\sqrt{x^2+y^2}}=\frac{1\cdot\sqrt{x^2+y^2}-y\frac{y}{\sqrt{x^2+y^2}}}{x^2+y^2}\\ q_x&=\frac{\partial}{\partial x}\frac{-x}{\sqrt{x^2+y^2}}=-\frac{1\cdot\sqrt{x^2+y^2}-x\frac{x}{\sqrt{x^2+y^2}}}{x^2+y^2}\end{aligned}\qquad \text{...no match!}$$

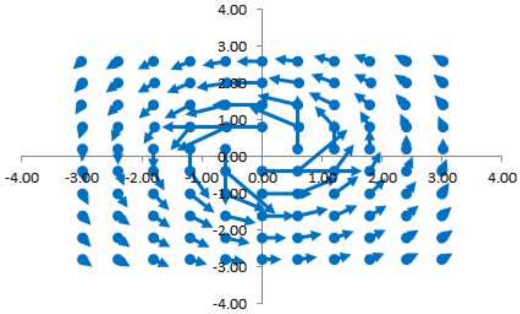
This vector field also fails the test and, therefore, isn't gradient.

Let's take this one step further:

$$W=\frac{V}{||V||^2}=\frac{1}{x^2+y^2}<y,\ -x>=\left\langle\frac{y}{x^2+y^2},\ -\frac{x}{x^2+y^2}\right\rangle=<p,q>\ .$$

The new vector field has the same directions but the magnitude varies; it approaches 0 as we move farther away from the origin and infinite as we approach the origin; i.e., we have:

$$W(X)\to 0\ \text{ as } ||X||\to\infty\quad\text{and}\quad ||W(X)||\to\infty\ \text{ as } X\to 0\,.$$



Let's test the condition:

$$\begin{aligned}p_y&=\frac{\partial}{\partial y}\frac{y}{x^2+y^2}=\frac{1\cdot(x^2+y^2)-y\cdot 2y}{(x^2+y^2)^2}=\frac{x^2-y^2}{(x^2+y^2)^2}\\ q_x&=\frac{\partial}{\partial x}\frac{-x}{x^2+y^2}=-\frac{1\cdot(x^2+y^2)-x\cdot 2x}{(x^2+y^2)^2}=-\frac{y^2-x^2}{(x^2+y^2)^2}\end{aligned}\qquad\text{...match!}$$

The vector field passes the test!

Does it mean that it is gradient then? No, it doesn't and we will demonstrate that it is not! The idea is the same that we started with in the beginning of the chapter: A round trip along the gradients is impossible as it leads to a net increase of the value of the function according to the *Monotonicity Theorem*. It is crucial that the vector field is undefined at the origin.

We will show later that the converse of the *Gradient Theorem* for dimension 2 isn't true:

$$p_y = q_x \not\Longrightarrow < p, q >= \nabla f \,,$$

unless a certain further restriction is placed. This restriction is *topological*: there can be no holes in the domain. Furthermore, integrating $p_y - q_x$ will be used to measure how close the vector field is to being gradient.

Chapter 5: The integral

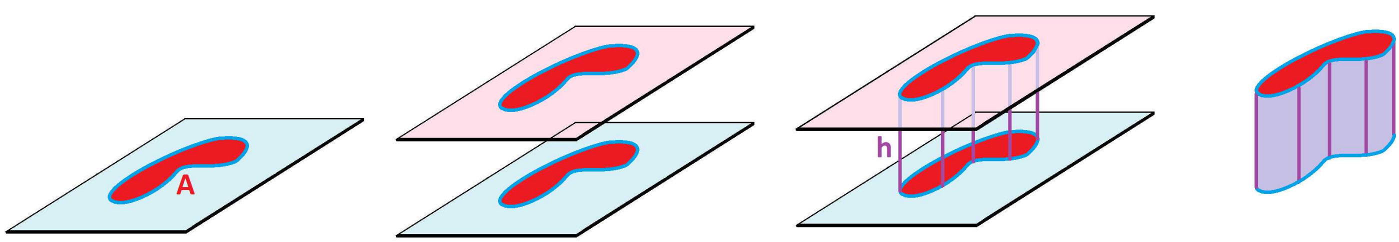
Contents

5.1. *Volumes and the Riemann sums*

All functions in this chapter are *real-valued*.

Our understanding of volumes is limited to that of Volume 3 (Chapter 3IC-3). If D is a region on the plane and it is lifted off the plane to the height h then the cylinder-like solid (a "shell") between these two plane regions is *assumed* to have the volume of:

$$V = A \cdot h\,.$$



We furthermore represented more complex solid in terms of these shells.

We will have to start over though.

Example 5.1.1: circle

Let's review the *Area Problem*. We confirm that the area of a circle of radius 1 is $A = \pi$. First we plot the graph of

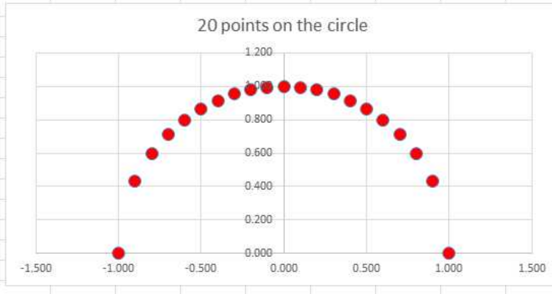
$$y = f(x) = \sqrt{1 - x^2},\ -1 \le x \le 1\,,$$

with 21 points (20 intervals). We let the values of x run from -1 to 1 every 0.1 and apply the formula:

```
=SQRT(1-RC3^2)
```

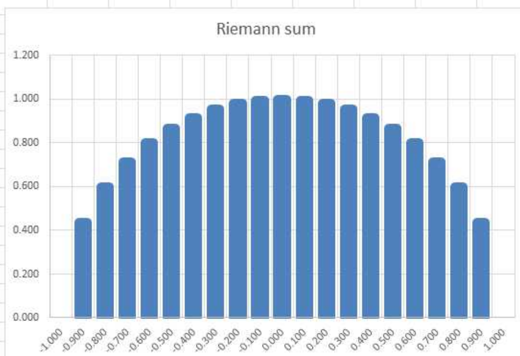
We get the values of y:

x	y
-1.000	0.000
-0.900	0.436
-0.800	0.600
-0.700	0.714
-0.600	0.800
-0.500	0.866
-0.400	0.917
-0.300	0.954
-0.200	0.980
-0.100	0.995
0.000	1.000
0.100	0.995
0.200	0.980
0.300	0.954
0.400	0.917
0.500	0.866
0.600	0.800
0.700	0.714
0.800	0.600



We next cover this half-circle with vertical bars based on the interval $[-1,1]$: the bases of the bars are our intervals in the x-axis and the heights are values of $y=f(x)$. To see the bars, we simply change the type of the chart plotted by the spreadsheet:

n=	20	intervals			
h=	0.10		Area=	1.552	1.571
	x	widths	heights	areas	
	-1.000		0.000		
	-0.900	0.100	0.436	0.044	
	-0.800	0.100	0.600	0.060	
	-0.700	0.100	0.714	0.071	
	-0.600	0.100	0.800	0.080	
	-0.500	0.100	0.866	0.087	
	-0.400	0.100	0.917	0.092	
	-0.300	0.100	0.954	0.095	
	-0.200	0.100	0.980	0.098	
	-0.100	0.100	0.995	0.099	
	0.000	0.100	1.000	0.100	
	0.100	0.100	0.995	0.099	
	0.200	0.100	0.980	0.098	
	0.300	0.100	0.954	0.095	
	0.400	0.100	0.917	0.092	
	0.500	0.100	0.866	0.087	
	0.600	0.100	0.800	0.080	
	0.700	0.100	0.714	0.071	



Then the area of the circle is approximated by the sum of the areas of the bars: we multiply the widths of the bars by the heights, place the result in the last column, and finally add all entries in this column. The result 1.552 is close to $\pi/2 \approx 1.571$.

We proceed to the *Volume Problem*. We will confirm that the volume of a sphere of radius 1 is $V=\frac{4}{3}\pi$. First we plot the graph of

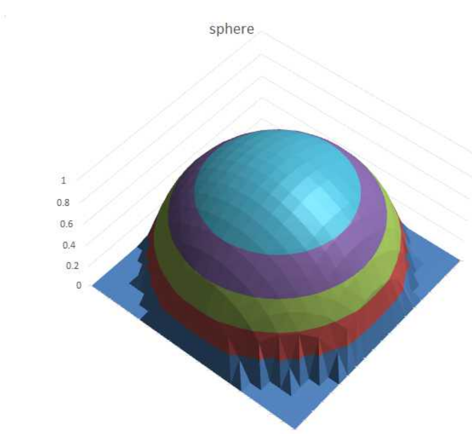
$$z=f(x,y)=\sqrt{1-x^2-y^2},\quad -1\le x\le 1,\ -1\le y\le 1\,.$$

We recycle our spreadsheet for the sphere. We already have 20 intervals for x in the first column. Now, just as before, we construct 20 intervals for y in the first row. We let the values of x and y run from -1 to 1 every 0.1 and apply the formula:

```
=SQRT(1-RC3^2-R4C^2)
```

to get the values of z:

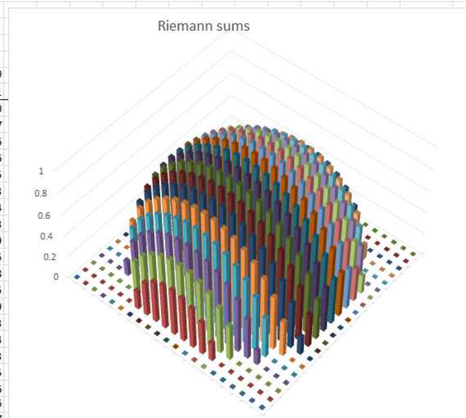
	y	-1.0	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
x	dx \ dy			0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
-1.00		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.90	0.1	0	0	0	0	0	0	0.173	0.316	0.387	0.424	0.436	0.424	0.387	0.316	0.173	0	0	0	0	0	0
-0.80	0.1	0	0	0	0	0	0.332	0.447	0.52	0.566	0.592	0.6	0.592	0.566	0.52	0.447	0.332	7E-09	0	0	0	0
-0.70	0.1	0	0	0	0.141	0.387	0.51	0.592	0.648	0.686	0.707	0.714	0.707	0.686	0.648	0.592	0.51	0.387	0.141	0	0	0
-0.60	0.1	0	0	0	0.387	0.529	0.624	0.693	0.742	0.775	0.794	0.8	0.794	0.775	0.742	0.693	0.624	0.529	0.387	1E-08	0	0
-0.50	0.1	0	0	0.332	0.51	0.624	0.707	0.768	0.812	0.843	0.86	0.866	0.86	0.843	0.812	0.768	0.707	0.624	0.51	0.332	0	0
-0.40	0.1	0	0.173	0.447	0.592	0.693	0.768	0.825	0.866	0.894	0.911	0.917	0.911	0.894	0.866	0.825	0.768	0.693	0.592	0.447	0.173	0
-0.30	0.1	0	0.316	0.52	0.648	0.742	0.812	0.866	0.906	0.933	0.949	0.954	0.949	0.933	0.906	0.866	0.812	0.742	0.648	0.52	0.316	0
-0.20	0.1	0	0.387	0.566	0.686	0.775	0.843	0.894	0.933	0.959	0.975	0.98	0.975	0.959	0.933	0.894	0.843	0.775	0.686	0.566	0.387	0
-0.10	0.1	0	0.424	0.592	0.707	0.794	0.86	0.911	0.949	0.975	0.99	0.995	0.99	0.975	0.949	0.911	0.86	0.794	0.707	0.592	0.424	0
0.00	0.1	0	0.436	0.6	0.714	0.8	0.866	0.917	0.954	0.98	0.995	1	0.995	0.98	0.954	0.917	0.866	0.8	0.714	0.6	0.436	0
0.10	0.1	0	0.424	0.592	0.707	0.794	0.86	0.911	0.949	0.975	0.99	0.995	0.99	0.975	0.949	0.911	0.86	0.794	0.707	0.592	0.424	0
0.20	0.1	0	0.387	0.566	0.686	0.775	0.843	0.894	0.933	0.959	0.975	0.98	0.975	0.959	0.933	0.894	0.843	0.775	0.686	0.566	0.387	0
0.30	0.1	0	0.316	0.52	0.648	0.742	0.812	0.866	0.906	0.933	0.949	0.954	0.949	0.933	0.906	0.866	0.812	0.742	0.648	0.52	0.316	0
0.40	0.1	0	0.173	0.447	0.592	0.693	0.768	0.825	0.866	0.894	0.911	0.917	0.911	0.894	0.866	0.825	0.768	0.693	0.592	0.447	0.173	0
0.50	0.1	0	0	0.332	0.51	0.624	0.707	0.768	0.812	0.843	0.86	0.866	0.86	0.843	0.812	0.768	0.707	0.624	0.51	0.332		0
0.60	0.1	0	0	0	0.387	0.529	0.624	0.693	0.742	0.775	0.794	0.8	0.794	0.775	0.742	0.693	0.624	0.529	0.387	2E-08		0
0.70	0.1	0	0	0	0.141	0.387	0.51	0.592	0.648	0.686	0.707	0.714	0.707	0.686	0.648	0.592	0.51	0.387	0.141			0
0.80	0.1	0	0	0		1E-08	0.332	0.447	0.52	0.566	0.592	0.6	0.592	0.566	0.52	0.447	0.332	2E-08				0
0.90	0.1	0	0	0				0.173	0.316	0.387	0.424	0.436	0.424	0.387	0.316	0.173						0
1.00	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



We next fill this half-sphere with vertical bars based on the square $[-1,1]\times[-1,1]$: the bases of the

bars (pillars) are our little squares in the xy-plane and the heights are values of z. To see the bars, we simply change the type of the chart plotted by the spreadsheet:

a	b		RS=	2.08069	2.094							
-1	1											
n=	20											
h=	0.1											
			y	-1.00	-0.90	-0.80	-0.70	-0.60	-0.50	-0.40	-0.30	-0.20
		x	dx \ dy			0.1	0.1	0.1	0.1	0.1	0.1	0.1
		-1.00		0	0	0	0	0	0	0	0	0
		-0.90	0.1	0	0	0	0	0	0	0.173	0.316	0.387
		-0.80	0.1	0	0	0	0	0	0.332	0.447	0.52	0.566
		-0.70	0.1	0	0	0	0.141	0.387	0.51	0.592	0.648	0.686
		-0.60	0.1	0	0	0	0.387	0.529	0.624	0.693	0.742	0.775
		-0.50	0.1	0	0	0.332	0.51	0.624	0.707	0.768	0.812	0.843
		-0.40	0.1	0	0.173	0.447	0.592	0.693	0.768	0.825	0.866	0.894
		-0.30	0.1	0	0.316	0.52	0.648	0.742	0.812	0.866	0.906	0.933
		-0.20	0.1	0	0.387	0.566	0.686	0.775	0.843	0.894	0.933	0.959
		-0.10	0.1	0	0.424	0.592	0.707	0.794	0.86	0.911	0.949	0.975
		0.00	0.1	0	0.436	0.6	0.714	0.8	0.866	0.917	0.954	0.98
		0.10	0.1	0	0.424	0.592	0.707	0.794	0.86	0.911	0.949	0.975
		0.20	0.1	0	0.387	0.566	0.686	0.775	0.843	0.894	0.933	0.959
		0.30	0.1	0	0.316	0.52	0.648	0.742	0.812	0.866	0.906	0.933
		0.40	0.1	0	0.173	0.447	0.592	0.693	0.768	0.825	0.866	0.894
		0.50	0.1	0	0	0.332	0.51	0.624	0.707	0.768	0.812	0.843
		0.60	0.1	0	0	0	0.387	0.529	0.624	0.693	0.742	0.775
		0.70	0.1	0	0	0	0.141	0.387	0.51	0.592	0.648	0.686
		0.80	0.1	0	0	0		1E-08	0.332	0.447	0.52	0.566
		0.90	0.1	0	0	0				0.173	0.316	0.387



We can see each row of bars as an approximation of the *area* of a slice of the sphere, which is another circle...

The volume of the sphere is now approximated by the sum of the volumes of the bars. Each of these volumes is the product of

- The height of the bar in this rectangle equal to the value of the function and with the ones outside the domain replaced with 0s.
- The area of the base is equal to 0.01.

Their sum is simply the sum of these heights multiplied by 0.01. The result produced by the spreadsheet is the following:

$$\text{Approximate volume of the hemisphere} = 2.081\,.$$

It is close to the theoretical result:

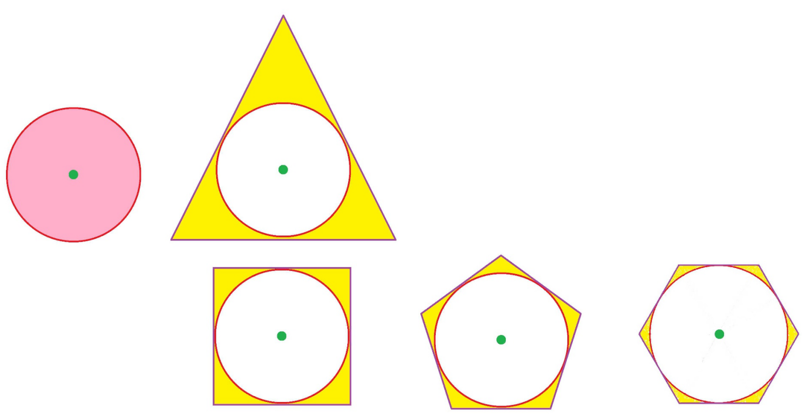
$$\text{Exact volume of the hemisphere} = 2\pi/3 \approx 2.094\,.$$

Exercise 5.1.2

Approximate the volume of the sphere radius 1 within 0.0001.

We have showed that indeed the area of a sphere of radius 1 is close to $A = \frac{4}{3}\pi$. But the real question is: What *is* the volume? One thing we do know. The volume of a box $a \times b \times c$ is abc. With that we can compute volume of various geometric figures with straight edges but what are the *volumes of curved objects*?

The idea, once again, comes from the ancient Greek's approach to understanding and computing the areas and volumes. They approximated the circle with regular polygons and the sphere with regular polyhedra:



The setup for the Riemann sums for functions of two variables is very similar to the one for numerical functions but by far more cumbersome.

Let's consider a rectangle $R = [a, b] \times [c, d],\ a < b,\ c < d$. Suppose also that we have two integers $n, m \geq 1$.

First, we have a partition of $[a, b]$ into n intervals of possibly different lengths:

$$[x_0, x_1],\ [x_1, x_2],\ ...,\ [x_{n-1}, x_n]\,,$$

with $x_0 = a,\ x_n = b$. The increments of x are:

$$\Delta x_i = x_i - x_{i-1},\ i = 1, 2, ..., n\,.$$

Second, we have a partition of $[c, d]$ into m intervals of possibly different lengths:

$$[y_0, y_1],\ [y_1, y_2],\ ...,\ [y_{m-1}, y_m]\,,$$

with $y_0 = c,\ y_m = d$. The increments of y are:

$$\Delta y_j = y_j - y_{j-1},\ j = 1, 2, ..., m\,.$$

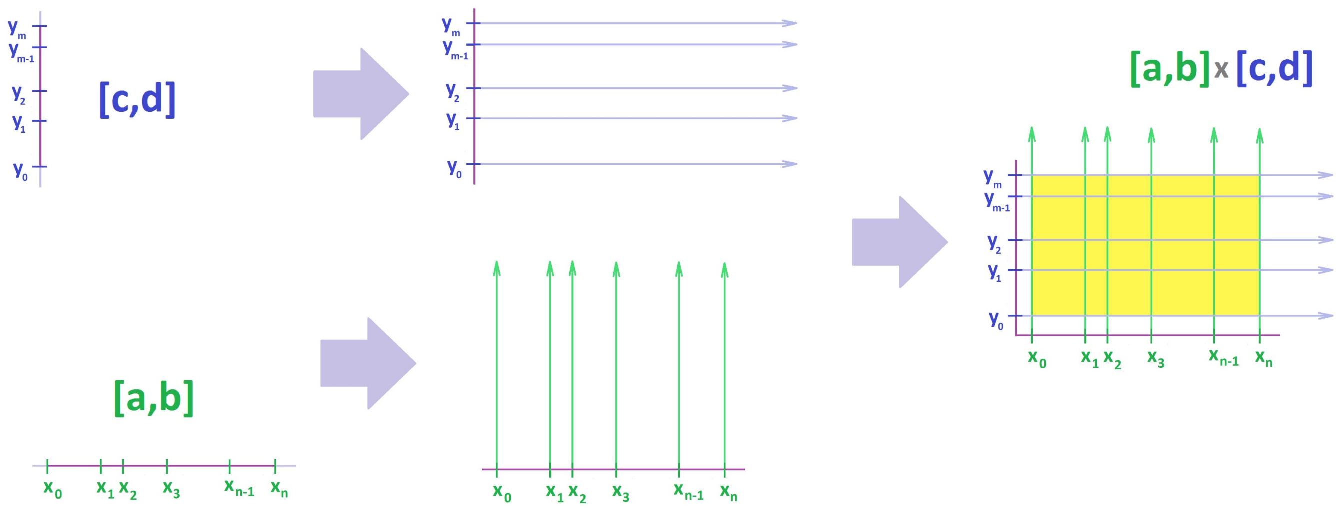
Altogether, we have a *partition* P of the rectangle $[a, b] \times [c, d]$ into smaller rectangles

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]\,.$$

These are 2-cells! The points

$$(x_i, y_j),\ i = 1, 2, ..., n,\ j = 1, 2, ..., m\,,$$

will be called the (primary) *nodes* of the partition:

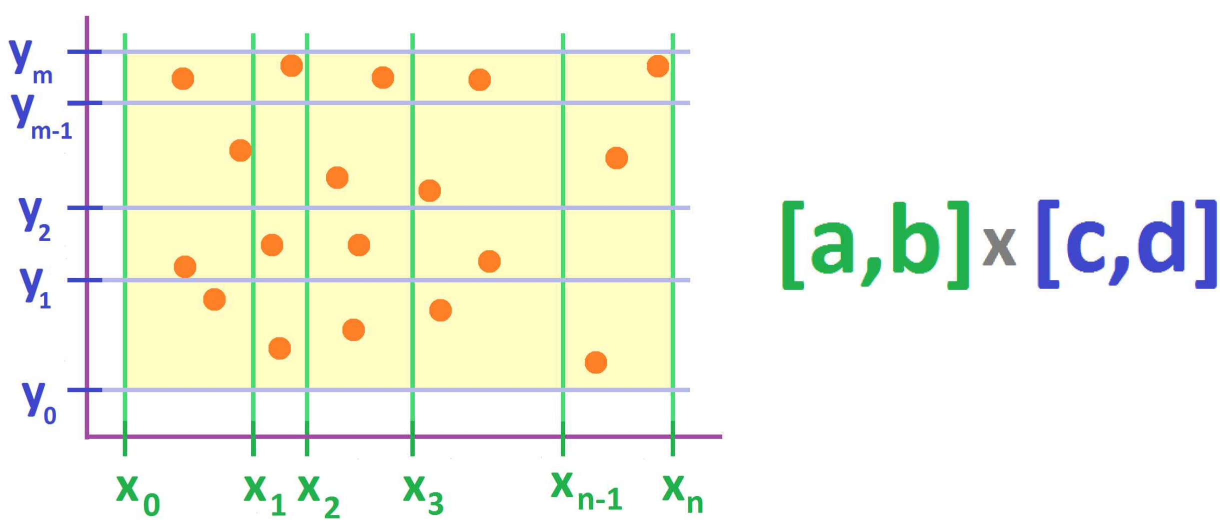


We won't need secondary nodes in this chapter.

We are also given the *tertiary nodes* of P for each pair $i = 0, 1, 2, ..., n-1$ and $i, j = 0, 1, 2, ..., m-1$:

- a point U_{ij} in the rectangle $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$.

Such a combination of rectangles and tertiary nodes in its intervals will be called an *augmented partition* P of R:



In the example above, the right upper corners were chosen.

Before we address how to compute volumes, let's consider a simpler problem.

Suppose a function $y = f(X) = f(x, y)$ defined at the tertiary nodes of the partition of the rectangle R and gives us the *amount* of some material contained in the corresponding cell. Then the total amount of the material in the whole rectangle is simply the sum of the values of f.

Definition 5.1.3: sum

The *sum* of a function $z = f(x, y)$ defined at the tertiary nodes of an augmented partition P of a rectangle $R = [a, b] \times [c, d]$ is defined to be the following:

$$\boxed{\sum_R f = \sum_{i=1}^{n} \sum_{j=1}^{m} f(U_{ij})}$$

Example 5.1.4: double summation

The double summation on the right follows the same idea as the single. For example, suppose we have these values a_{ij} arranged in an array:

$j \backslash i$	1	2	3
1	2	0	1
2	-1	2	3

Then,

$$\begin{aligned} \sum_{i=1}^{3} \sum_{j=1}^{2} a_{ij} &= \sum_{i=1}^{3} \left(\sum_{j=2}^{2} a_{ij} \right) \\ &= \sum_{i=1}^{3} \left(a_{i1} + a_{i2} \right) \\ &= (a_{11} + a_{12}) + (a_{21} + a_{22}) + (a_{31} + a_{32}) \\ &= (2 + (-1)) + (0 + 2) + (1 + 3) \\ &= 7\,. \end{aligned}$$

There is also a simple way to add all numbers in an array presented as a spreadsheet:

```
=sum(R1C1:R2C3)
```

Exercise 5.1.5

For the above table compute:

$$\sum_{j=1}^{2} \sum_{i=1}^{3} a_{ij}\,.$$

Note that when tertiary nodes aren't provided, we can think of the 2-cells themselves as the inputs of the function: $U_{ij} = R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$. This makes f a 2-form.

The area of each 2-cell is:

$$\Delta A_{ij} = \Delta x_i \cdot \Delta y_j\,.$$

In other words, the product of the increments of x and y is the increment of the area.

Suppose next we have a function $y = f(X) = f(x, y)$ defined at the tertiary nodes of the partition of the

rectangle R and gives us the *height* of a bar on top of the corresponding cell:

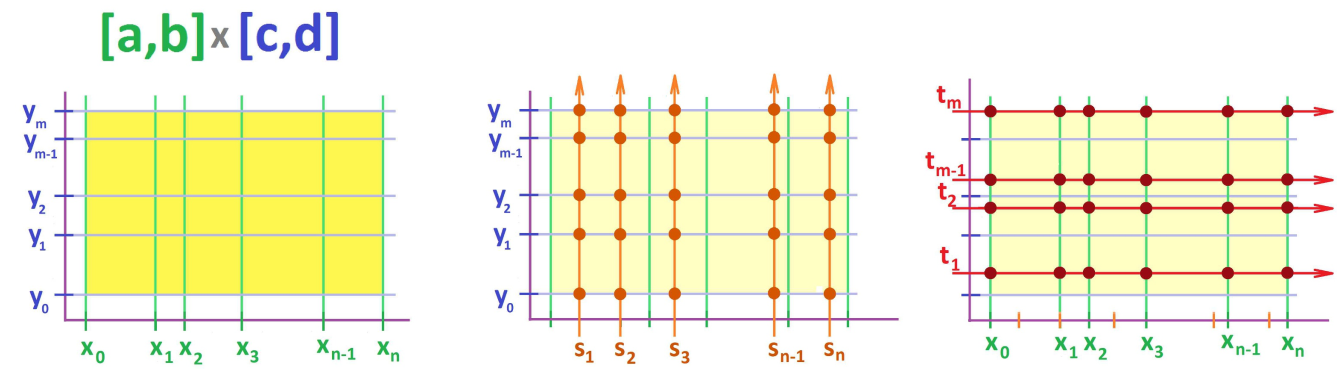
$$\text{the volume of } ij \text{ bar} = \underbrace{f(U_{ij})}_{\text{height of bar}} \cdot \overbrace{\Delta x_i}^{\text{depth of base}} \cdot \overbrace{\Delta y_j}^{\text{width of base}}$$

We then add all of these together in order to compute the *volume of the solid under the graph* of $z = f(x, y)$ over rectangle R.

Example 5.1.6: partitions

We consider the particular case when the tertiary nodes of P come from the secondary nodes of the augmented partitions of the intervals $[a, b]$ and $[c, d]$:

$$U_{ij} = (s_i, t_j)\,.$$



The computation is illustrated below:

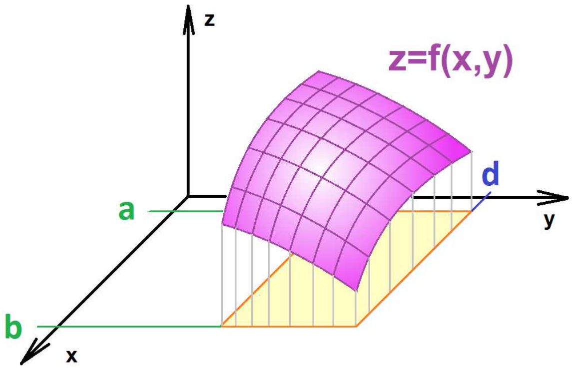
	y_0	Δy_1	y_1	Δy_2	y_3	...	y_{m-1}	Δy_m	y_m
x_0	•	−−	•	−−	•	...	•	−−	•
Δx_1	|	$f(s_1, t_1)\Delta x_1 \Delta y_1$	|	$f(s_1, t_2)\Delta x_1 \Delta y_2$	|	...	|	$f(s_1, t_m)\Delta x_1 \Delta y_m$	|
x_1	•	−−	•	−−	•	...	•	−−	•
Δx_2	|	$f(s_2, t_1)\Delta x_2 \Delta y_1$	|	$f(s_2, t_2)\Delta x_2 \Delta y_2$	|	...	|	$f(s_2, t_m)\Delta x_2 \Delta y_m$	|
x_2	•	−−	•	−−	•	...	•	−−	•
..	.	..	.	..	.	...	.	..	.
x_{n-1}	•	−−	•	−−	•	...	•	−−	•
Δx_n	|	$f(s_n, t_1)\Delta x_n \Delta y_1$	|	$f(s_n, t_2)\Delta x_n \Delta y_2$	|	...	|	$f(s_n, t_m)\Delta x_n \Delta y_m$	|
x_n	•	−−	•	−−	•	...	•	−−	•

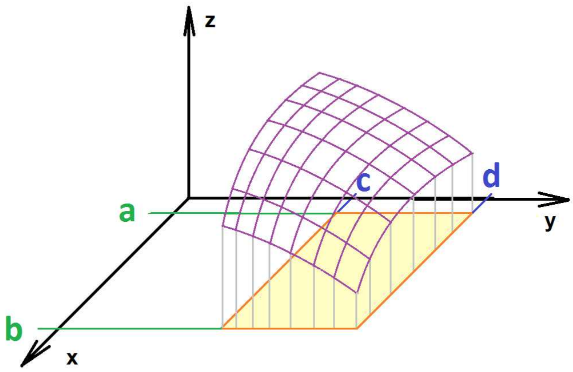
Definition 5.1.7: Riemann sum

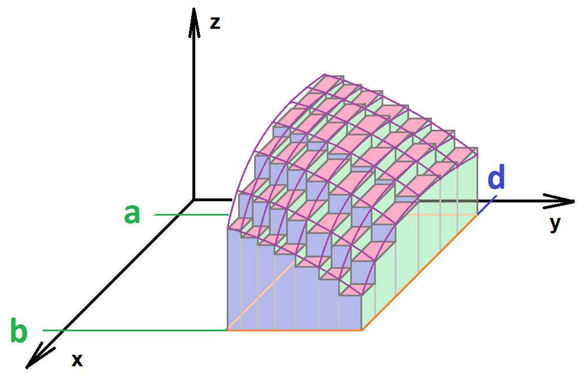
The *Riemann sum* of a function $z = f(x, y)$ defined at the tertiary nodes of an augmented partition P of a rectangle $R = [a, b] \times [c, d]$ is defined and denoted to be the following:

$$\boxed{\sum_R f(U_{ij})\, \Delta x_i \Delta y_j = \sum_{i=1}^{n} \sum_{j=1}^{m} f(U_{ij}) \Delta x_i \Delta y_j}$$

The Riemann sum of a sampled function of two variables is shown below:







The abbreviated notation for the Riemann sum is:

Riemann sum

$$\sum_R f\,\Delta x \Delta y$$

Example 5.1.8: plane

Let's consider a very simple example:

$$f(x,y) = x + y,\ \ R = [0,1] \times [0,1]\,.$$

We choose

$$n = 2,\ \ m = 2\,.$$

Then the end-points of the intervals are:

$$x_0 = 0,\ \ x_1 = .5,\ \ x_2 = 1 \quad \text{and} \quad y_0 = 0,\ \ y_1 = .5,\ \ y_2 = 1\,.$$

The nodes are

$$\begin{array}{ccccc} (0,1) & - & (.5,1) & - & (1,1) \\ | & & | & & | \\ (0,.5) & - & (.5,.5) & - & (1,.5) \\ | & & | & & | \\ (0,0) & - & (.5,0) & - & (1,0) \end{array}$$

They are the corners of the rectangles of the partition:

$$[0,.5] \times [.5,1] \quad [.5,1] \times [.5,1] [0,.5] \times [0,.5] \quad [.5,1] \times [0,.5]$$

Now we choose the tertiary nodes. Specifically, let's choose the bottom left corners:

$(0,.5)$	$(.5,.5)$
$(0,0)$	$(.5,0)$

leading to

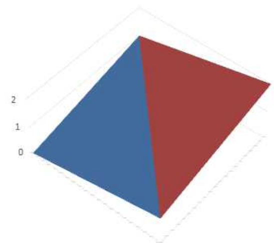
$f(0,.5) = .5$	$f(.5,.5) = 1$
$f(0,0) = 0$	$f(.5,0) = .5$

leading to

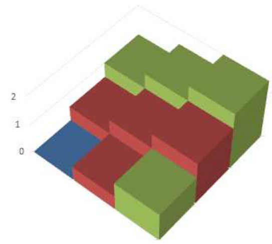
$.5 \cdot .5^2 = .125$	$1 \cdot .5^2 = .25$
$0 \cdot .5^2 = 0$	$.5 \cdot .5^2 = .125$

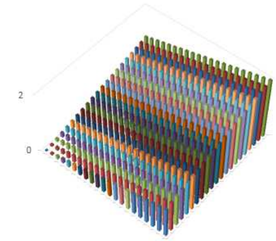
The values of the function and then the volumes of the bars are shown on right. Then the sum of those is:

$$\sum_R f\,\Delta x\Delta y = .5\,.$$



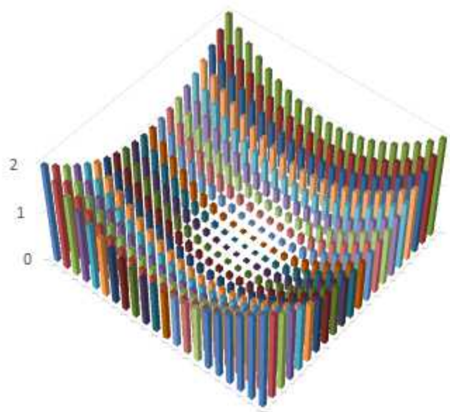
			RS=	2.25			
0	1						
n=	2						
h=	0.5		y	0	0.5	1	
		x	widths	heights			
		0		0	0.5	1	
		0.5	0.5	0.5	1	1.5	
		1	0.5	1	1.5	2	





Example 5.1.9: paraboloid

Riemann sums of the paraboloid of revolution:



Just as in the one-dimensional case, we are allowed to have negative values of f, with possibly negative volumes of the bars. These are the signed distance and the *signed volume* respectively. We speak then of the volume of the solid *between* the graph of $z = f(x, y)$ and the rectangle R in the xy-plane.

Furthermore, we can have negative lengths for the independent variables too. Suppose $a < b$ and $c < d$, then

- The rectangles $[a, b] \times [c, d]$ and $[b, a] \times [d, c]$ are *positively oriented* and have positive areas.
- The rectangles $[a, b] \times [d, c]$ and $[b, a] \times [c, d]$ are *negatively oriented* and have negative areas.

Once again, the Riemann sum over an oriented rectangle is referred to as the *signed volume.*

The algebraic properties of sum and Riemann sums are similar to the one in Volume 3 (Chapter 3IC-1).

Theorem 5.1.10: Constant Function Rule

Suppose f is constant on a rectangle R, i.e., $f(x, y) = m$ for all (x, y) in R and some real number c. Then

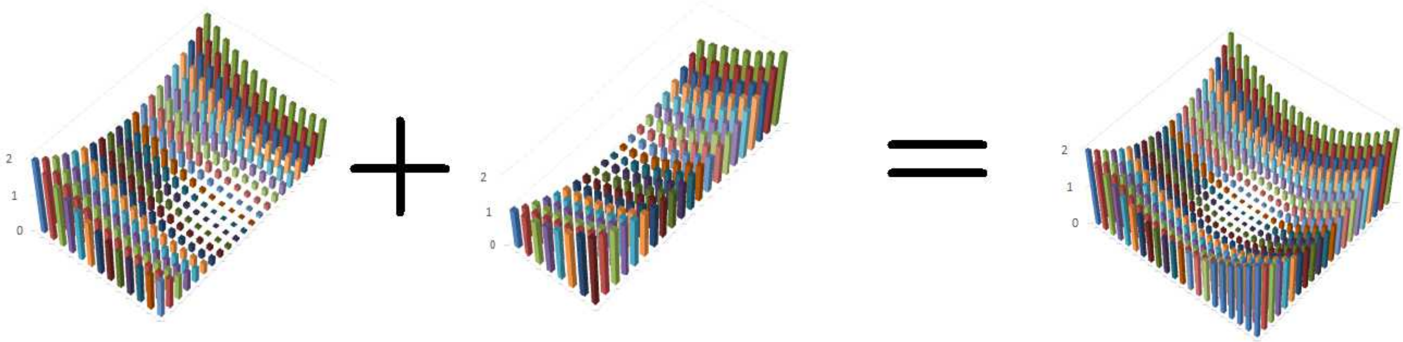
$$\boxed{\sum_R f\,\Delta x\Delta y = m(b-a)(d-c)}$$

Indeed, the Riemann sum represents the area of the rectangle with width $b - a$, the depth $d - c$, and the height m.

5.2. Properties of the Riemann sums

In spite of their somewhat complex definition, the Riemann sums are finite! There are no issues of infinity or divergence to worry about. They are, therefore, subject to the *usual rules of algebra*: Their terms can be multiplied, factored, re-arranged and added in a different order, etc.

If we proceed to the adjacent rectangle (just as with intervals in Volume 3), we can just continue to add terms of the Riemann sum:



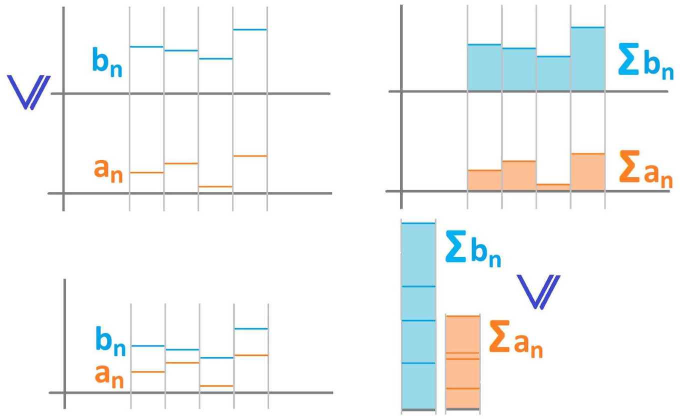
Theorem 5.2.1: Additivity Rule For Riemann Sums

Suppose $z = f(x, y)$ is a function. Suppose a, b, c, d, q are any numbers and suppose we have partitions of the intervals $[a, q]$, $[q, b]$, $[c, d]$ and augmented partitions of the rectangles $[a, q] \times [c, d]$ and $[q, b] \times [c, d]$. Then we have:

$$\boxed{\sum_{[a,q]\times[c,d]} f\,\Delta x\Delta y + \sum_{[q,b]\times[c,d]} f\,\Delta x\Delta y = \sum_{[a,b]\times[c,d]} f\,\Delta x\Delta y}$$

with summations over these partitions.

If two functions are comparable then so are the terms of their Riemann sums:



We have a similar result for volumes:

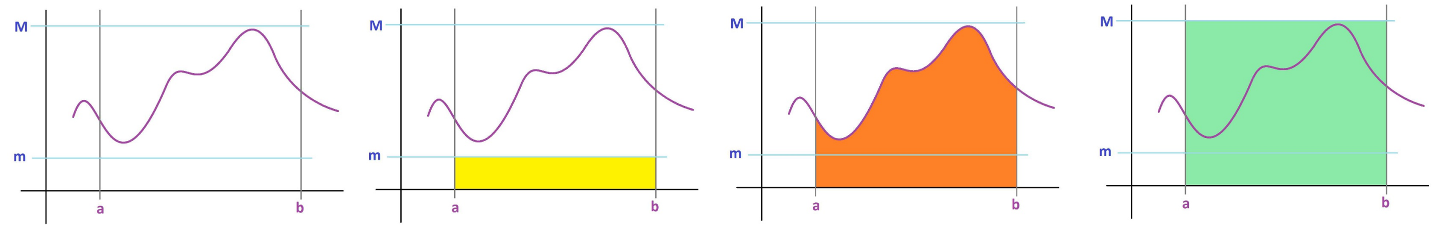
Theorem 5.2.2: Comparison Rule For Riemann Sums

Suppose f and g are functions defined on a rectangle R. Then we have:

$$(1)\ \ f(x, y) \ge g(x, y) \text{ on } R \implies \sum_R f\,\Delta x\Delta y \ge \sum_R g\,\Delta x\Delta y\,.$$

$$(2)\ f(x,y) < g(x,y) \text{ on } R \implies \sum_R f\,\Delta x\Delta y < \sum_R g\,\Delta x\Delta y\,.$$

Estimation of areas was considered in Volume 3:



We have a similar result for volumes:

Theorem 5.2.3: Estimate Rule For Riemann Sums

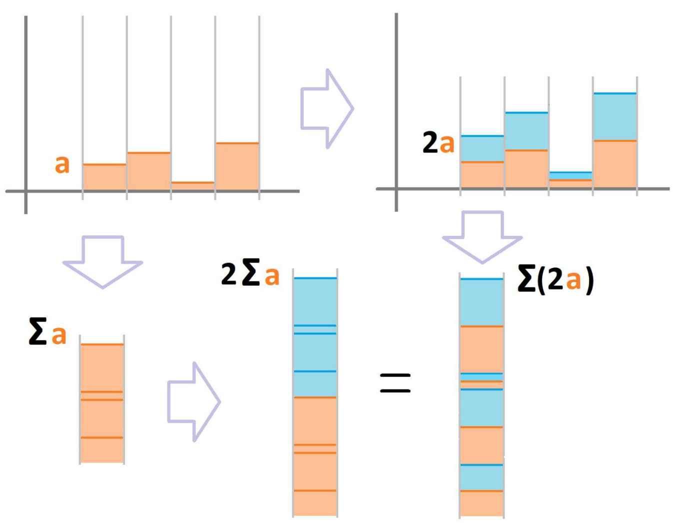
Suppose $z = f(x,y)$ is a function. For any a, b with $a < b$ and any c, d with $c < d$, if

$$m \le f(x,y) \le M\,,$$

for all x with $a \le x \le b$ and all y with $c \le y \le d$, then

$$m(b-a)(d-c) \le \sum_{[a,b]\times[c,d]} f\,\Delta x\Delta y \le M(b-a)(d-c)$$

The picture below illustrates the idea of multiplication of the function viz. multiplication of the volume under its graph:



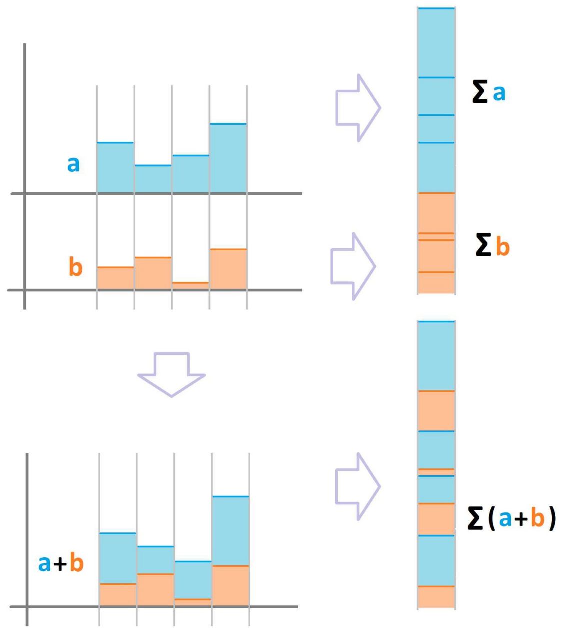
We have a similar result for addition over rectangles:

Theorem 5.2.4: Constant Multiple Rule For Riemann Sums

Suppose f is a function. For any rectangle R and any real c, we have:

$$\sum_R (c\cdot f)\,\Delta x\Delta y = c\cdot\sum_R f\,\Delta x\Delta y$$

The picture below illustrates the idea of adding functions viz. adding the volumes under their graphs:



Theorem 5.2.5: Sum Rule For Riemann Sums

Suppose f and g are functions. For any rectangle R, we have:

$$\boxed{\sum_R (f+g)\,\Delta x\Delta y = \sum_R f\,\Delta x\Delta y + \sum_R g\,\Delta x\Delta y}$$

Proof.

This is a simple algebraic manipulation of the corresponding Riemann sums:

$$\begin{aligned}\sum_R (f+g)\,\Delta x\Delta y &= \sum_{ij} \left(f(s_i,t_j)+g(s_i,t_j)\right)\Delta A_{ij}\\ &= \sum_{ij}\left(f(s_i,t_j)\Delta A_{ij}+g(s_i,t_j)\Delta A_{ij}\right)\\ &= \sum_{ij} f(s_i,t_j)\Delta A_{ij}+\sum_{ij} g(s_i,t_j)\Delta A_{ij}\\ &= \sum_R f\,\Delta x\Delta y+\sum_R g\,\Delta x\Delta y\,.\end{aligned}$$

Exercise 5.2.6

Prove the rest of these theorems.

The summation of the Riemann sum can be carried out in two main ways. We can add the rows first and then the resulting column or vice versa. A more profound way is to recognize the presence of functions of single variable in this function of two variables as well as the Riemann sums of these functions in its Riemann sum.

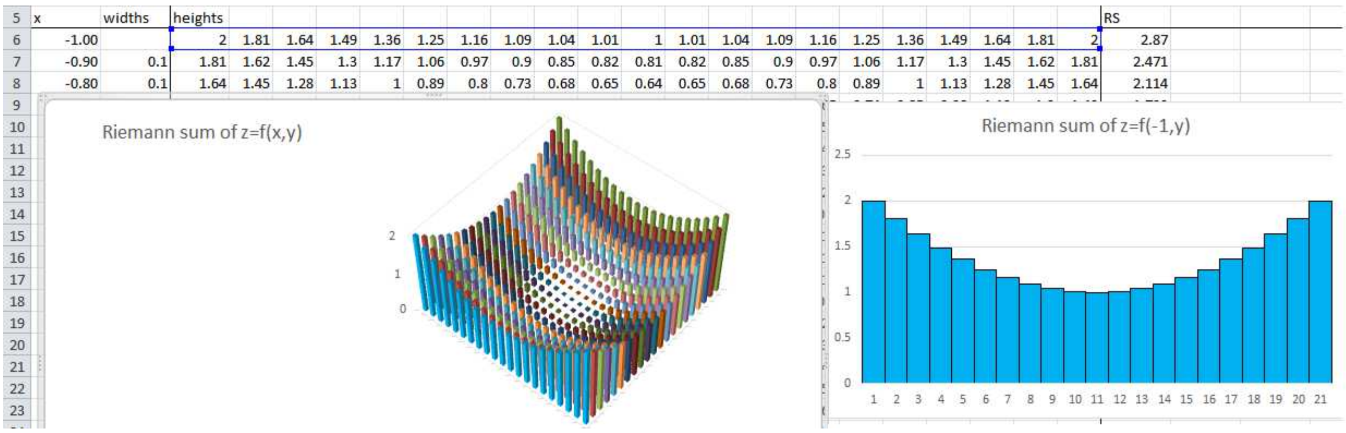
Fixing any value of x creates a new single variable function of y from $z=f(x,y)$:

$$f_x(y)=f(x,y)\,.$$

We pick one of the end-points $x=x_i$ of the partition of $[a,b]$ and form the Riemann sum of this function (with respect to y over $[c,d]$):

$$\sum_c^d f_{x_i}\,\Delta y=\sum_{j=1}^m f(s_i,t_j)\Delta y_j\,.$$

For the spreadsheet, the Riemann sum is computed and placed at the cell at its end of each row. As a result the highlighted row of bars in the Riemann sum of the function of two variables, a and y, reappears as the Riemann sum of a function of single variable, y:



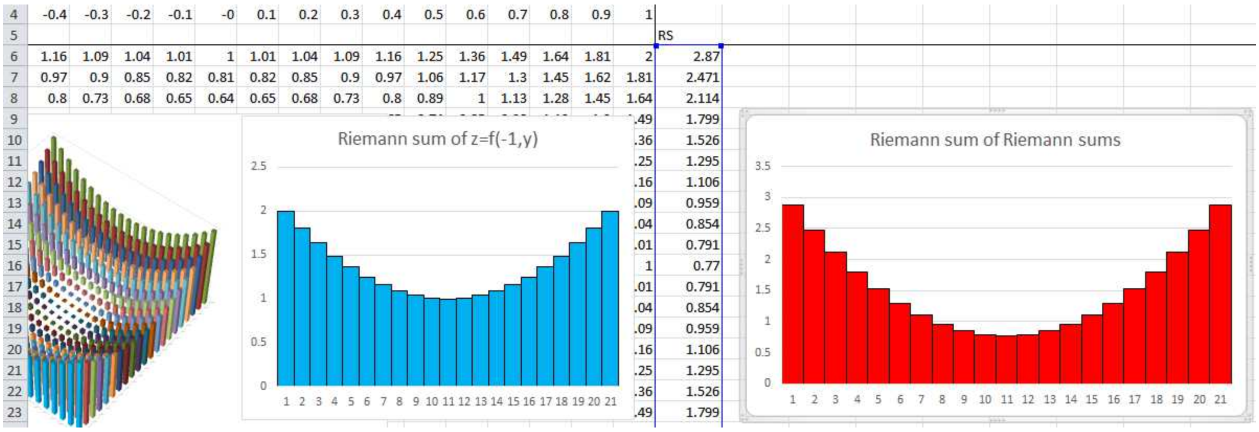
When this is done for each x_i, the result is a new function g defined on the end-points of the partition of $[c, d]$:

$$g(x_i) = \sum_{c}^{d} f_{x_i}\,\Delta y\,.$$

We now compute its Riemann sum (with respect to x over $[a, b]$):

$$\sum_R f\,\Delta x \Delta y = \sum_{i=1}^{n} g(x_i)\Delta x_i\,,$$

producing the Riemann sum of the original function f. We show this computation for a spreadsheet below:



For the spreadsheet, the Riemann sums placed at the cell at the end of each row, together, produce a column of values, i.e., a function of single variable, x. Its Riemann sum is now computed and placed at its bottom. The result is the same (of course!) number as the original, "double" Riemann sum.

Alternatively, for each column, the Riemann sum is computed and placed at the cell at its bottom. Together, the numbers give us a row of values. Its Riemann sum is now computed and placed at its end. The results are the same.

Exercise 5.2.7

Carry out the missing part of the above analysis starting with: "Fixing any value of y creates a new single variable function of x from $z = f(x, y)$...".

Our analysis amounts to the following formula:

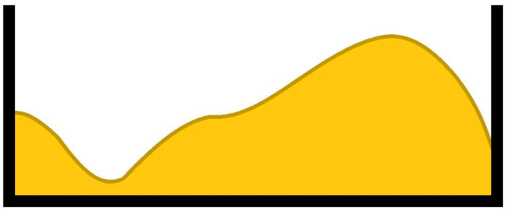
Theorem 5.2.8: Riemann Sum of Riemann Sums

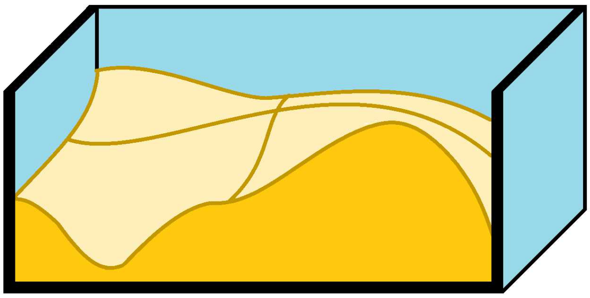
$$\begin{aligned}\sum_R f\,\Delta x\Delta y &= \sum_{i=1}^{n}\left(\sum_{j=1}^{m} f(s_i,t_j)\Delta y_j\right)\Delta x_i\\ &= \sum_{j=1}^{m}\left(\sum_{i=1}^{n} f(s_i,t_j)\Delta x_i\right)\Delta y_j\end{aligned}$$

In other words, the order of summation doesn't matter.

5.3. The Riemann integral over rectangles

The familiar integral of a numerical function is the (algebraic) *area under the graph* and now this is the *volume under the graph* of a function of two variables. It is as if we used to look at a certain sandbox from aside and tried to find the area of the visible sand as a way to determine the amount of sand in it:





Now we stand up and see that this was just a cross-section and we need to start to think three-dimensional...

Recall that the function is defined on the rectangle partitioned into smaller rectangles and then it is *sampled* – in a possibly non-uniform manner – in each. The result is a two-dimensional *array* of numbers:

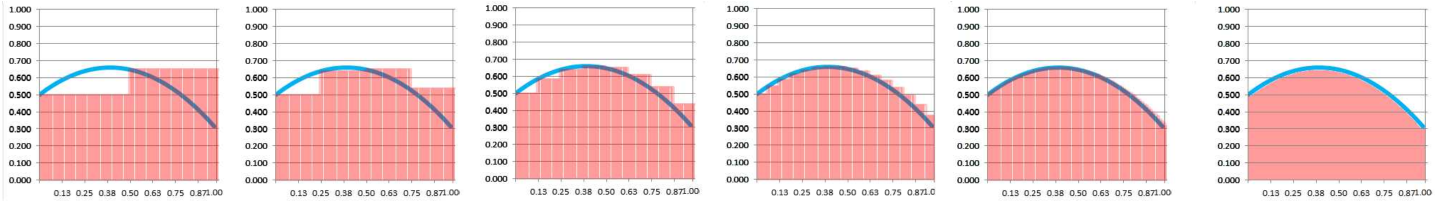
$$\text{heights:}\ \begin{array}{|ccc} \hline f(s_1,t_1) & f(s_1,t_2) & ...\\ f(s_2,t_1) & f(s_2,t_2) & ...\\ ... & ... & ... \end{array} \quad \to \quad \text{volumes:}\ \begin{array}{|ccc} \hline f(s_1,t_1)\Delta A_{11} & f(s_1,t_2)\Delta A_{12} & ...\\ f(s_2,t_1)\Delta A_{21} & f(s_2,t_2)\Delta A_{22} & ...\\ ... & ... & ... \end{array}$$

The sum of all the members of this array is the Riemann sum of the function.

Now, the Riemann sums are just approximations of the volume and in order to improve them, we have to *refine* the partition. And we keep refining, so that we have simultaneously:

$$n,m\to\infty \ \text{ and } \ \Delta x_i,\Delta y_j\to 0\,.$$

So, this is wat is happening for *both* x and y:



To make this idea specific, we define the *mesh of a partition* P as follows:

$$|P| = \max_{i,j} \{\Delta x_i,\ \Delta y_j\}\,.$$

It is a measure of "refinement" of P.

Definition 5.3.1: Riemann integral

The *Riemann integral* of a function $z = f(x, y)$ over rectangle $R = [a, b] \times [c, d]$ is defined to be the limit of a sequence of its Riemann sums with the mesh of their augmented partitions P_k approaching 0 as $k \to \infty$. When all these limits exist and are all equal to each other, f is called an *integrable function over* R and the result is denoted as follows:

$$\boxed{\iint_R f(x,y)\,dxdy = \lim_{k\to\infty} \sum_R f_k\,\Delta x \Delta y}$$

where f_k is f sampled at the secondary nodes of the partition. It is also called the *definite integral*, the *double integral*, or simply *integral*, of f over R called the *domain of integration*. When all these limits are equal to $+\infty$ (or $-\infty$), we say that the integral is *infinite* and write:

$$\iint_R f(x,y)\,dxdy = +\infty \text{ (or } -\infty)\,.$$

Warning!

There is no integration over x or y here.

An abbreviated notation is as follows:

Double integral

$$\iint_R f\,dA$$

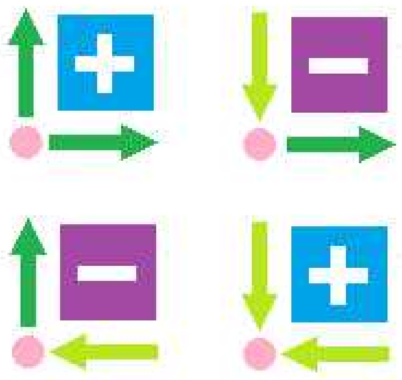
Here "A" stands for "area".

We will also refer to the integral as well as the (signed) *volume under the graph* of f or the (signed) *volume between the graph and the xy-plane.*

The custom of placing the domain of integration under the integral sign is also applied to the numerical functions, as follows:

$$\int_a^b f\,dx = \int_{[a,b]} f\,dx\,.$$

Instead of "from a to b" we have now an integral "over $[a, b]$". The advantage of the old notation is that it specifies the *orientation of the segment* $[a, b]$. From now on, all orientations are positive by default.



In general, we follow the rule:

$$\text{orientation of } [a,b] = \operatorname{sign}(b-a)\,,$$

and

$$\text{orientation of } [a,b]\times[c,d] \; = \operatorname{sign}((b-a)(d-c))\,.$$

Let's verify the definition for a simple function.

Theorem 5.3.2: Constant Integral Rule

Suppose $z = f(x,y)$ is constant on rectangle $R = [a,b]\times[c,d]$, i.e., $f(x,y) = m$ for all (x,y) in R and some real number m. Then f is integrable on R and

$$\boxed{\iint_R f\,dA = m(b-a)(d-c)}$$

The following result proves that our definition makes sense for a large class of functions.

Theorem 5.3.3: Cont. => Integr.

All continuous functions on rectangle R are integrable on R.

Proof.

The converse isn't true because the sign function $\operatorname{sign}(xy)$ is integrable over any rectangle.

We once again utilize the idea of signed distance, signed area, and signed volume.

Theorem 5.3.4: Orientation of Domain of Integration

The Riemann integral of a function $z = f(x,y)$ over rectangle $[b,a]\times[c,d]$ is equal to the negative of the integral over $[a,b]\times[c,d]$:

$$\iint_{[b,a]\times[c,d]} f\,dA = -\iint_{[a,b]\times[c,d]} f\,dA\ ;$$

and similarly,

$$\iint_{[a,b]\times[d,c]} f\,dA = -\iint_{[a,b]\times[c,d]} f\,dA\,;$$

but

$$\iint_{[b,a]\times[d,c]} f\,dA = \iint_{[a,b]\times[c,d]} f\,dA\,.$$

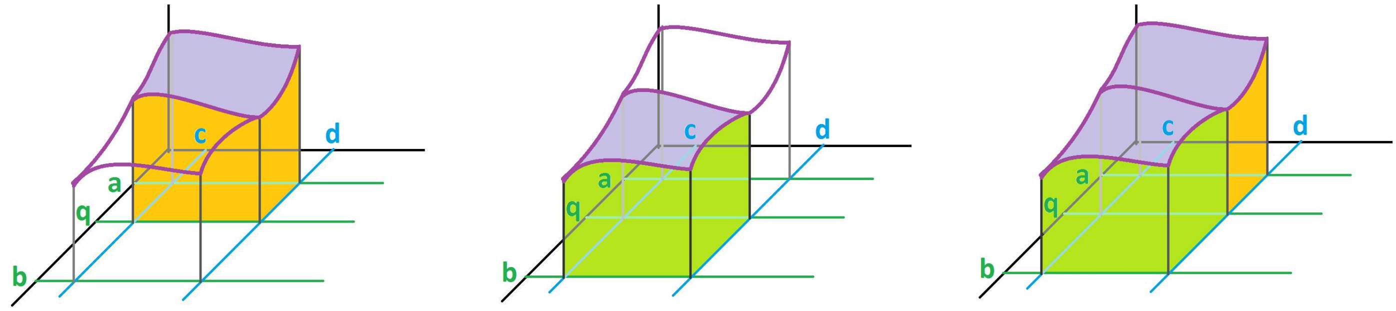
It all depends on whether the rectangle is positively or negatively oriented!

Now the properties of the Riemann integral. The mimic the ones for numerical functions but they follow from the corresponding properties of the Riemann sums (and the analogous rules of limits).

The additivity property for integrals – combining the domains of integration – of numerical functions,

$$\int_a^b f(x)\,dx + \int_a^b f(x)\,dx = \int_a^b f(x)\,dx\,,$$

has a match. The interpretation is the same: the amount of a quantity contained in a region formed by two regions with a negligible overlap is equal to the sum of the quantities in the two. We just look at the region from above which reveals the third dimension. So, we switch from areas to volumes:



It's as if we put a divider in our sandbox...

The result below follows from the *Additivity for Riemann Sums*.

Theorem 5.3.5: Additivity of Double Integral

Suppose a function $z = f(x, y)$ is integrable over the rectangles

$$R = [a, q] \times [c, d] \quad \text{and} \quad S = [q, b] \times [c, d]\,.$$

Then f is integrable over the rectangle $R \cup S = [a, b] \times [c, d]$ and we have:

$$\boxed{\iint_R f\,dA + \iint_S f\,dA = \iint_{R\cup S} f\,dA}$$

Thus, adding the domains of integration adds the integrals too.

Theorem 5.3.6: Local Integrability For Double Integrals

If $z = f(x, y)$ is integrable over R then it is also integrable over any rectangle R' in R.

The following is another important corollary.

Theorem 5.3.7: Cont. => Integr.

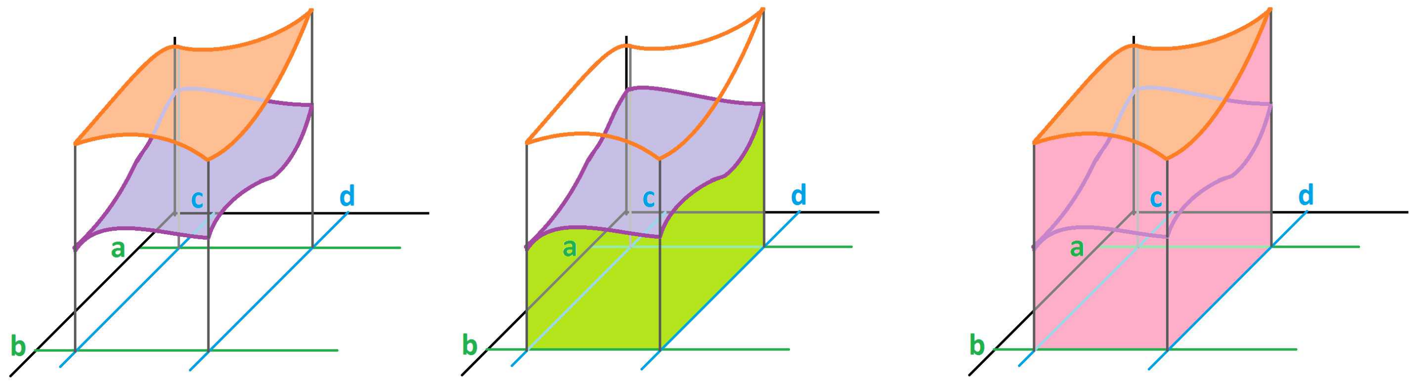
All piecewise continuous functions are integrable.

Proof.

It follows from the integrability of continuous functions and the Additivity Rule.

In particular, all step-functions are integrable.

Just as before, the larger function contains a larger volume under its graph:



Theorem 5.3.8: Comparison Rule For Double Integrals

If

$$f(x,y) \geq g(x,y) \text{ for all } (x,y) \text{ in rectangle } R\,,$$

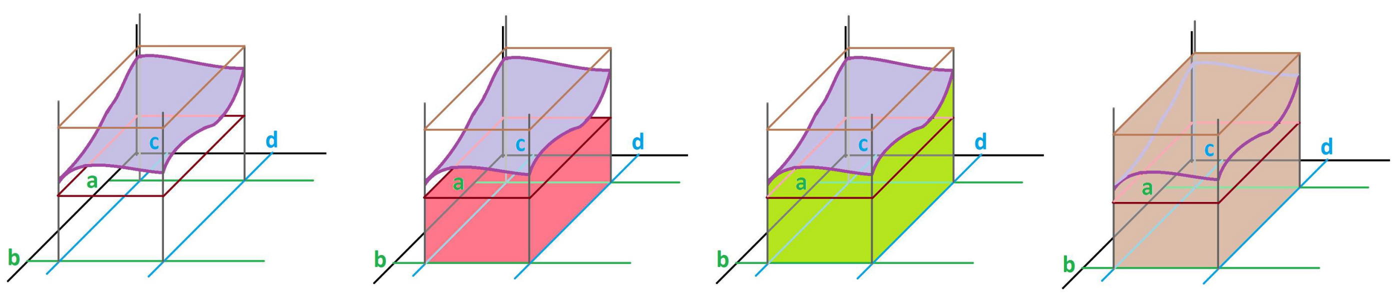
then

$$\boxed{\iint_R f\,dA \geq \iint_R g\,dA}$$

provided $z = f(x,y)$ *and* $z = g(x,y)$ *are integrable functions over* R*. Otherwise we have:*

$$\boxed{\begin{array}{ccc} \displaystyle\iint_R f\,dA = -\infty & \Longrightarrow & \displaystyle\iint_R g\,dA = -\infty \\ \displaystyle\iint_R f\,dA = +\infty & \Longleftarrow & \displaystyle\iint_R g\,dA = +\infty \end{array}}$$

If we know only estimates of the function, we have an estimate – below and above – for the volume under its graph.



For the general case, we have the following.

Theorem 5.3.9: Estimate Rule For Double Integrals

Suppose $z = f(x,y)$ *is an integrable function over* $R = [a,b] \times [c,d]$*. Then, if* $a < b$ *and* $c < d$ *and*

$$m \leq f(x,y) \leq M\,,$$

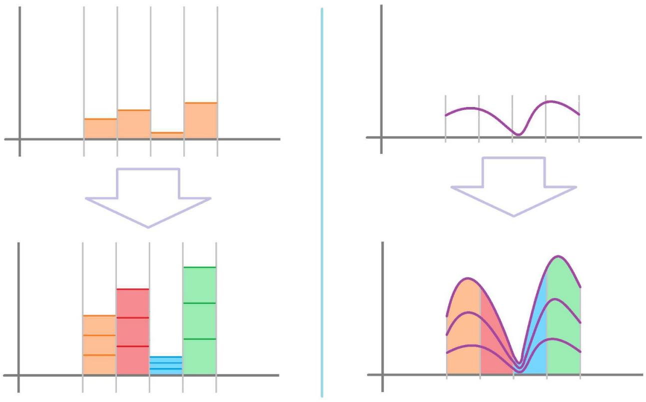
for all (x,y) *in* R*, we have:*

$$\boxed{m(b-a)(d-c) \leq \iint_R f\,dA \leq M(b-a)(d-c)}$$

Exercise 5.3.10

What if the orientation of the rectangle is negative?

Finally, these are the algebraic properties. First, the picture below illustrates the idea that tripling the height of a surface will need tripling the amount of soil under it:

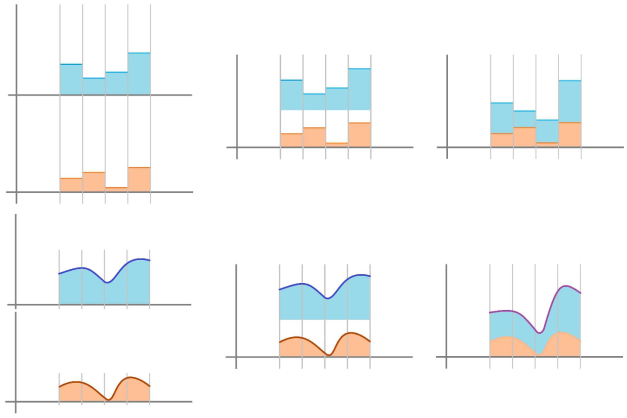


Theorem 5.3.11: Constant Multiple Rule For Double Integrals

Suppose $z = f(x,y)$ is an integrable function over a rectangle R. Then so is $c \cdot f$ for any real c and we have:

$$\boxed{\iint_R (c \cdot f)\, dA = c \cdot \iint_R f\, dA}$$

Finally, the picture below illustrates what happens when the bottom drops from a bucket of sand and it falls on a curved surface:



Theorem 5.3.12: Sum Rule For Double Integrals

Suppose $z = f(x,y)$ and $z = g(x,y)$ are integrable functions over a rectangle R. Then so is $f + g$ and we have:

$$\boxed{\iint_R (f+g)\, dA = \iint_R f\, dA + \iint_R g\, dA}$$

Proof.

We take the limit, as $|P| \to 0$, of the *Sum Rule for Riemann sums*:

$$\begin{array}{ccccc}
\sum_R (f+g)\,\Delta x \Delta y & = & \sum_R f\,\Delta x \Delta y & + & \sum_R g\,\Delta x \Delta y \\
\downarrow & & \downarrow & & \downarrow \\
\iint_R (f+g)\,dA & & \iint_R f\,dA & & \iint_R g\,dA
\end{array}$$

The transition is justified by the *Sum Rule for Limits.*

Exercise 5.3.13

Prove the rest of these theorems.

The actual computations of integrals require *single integrals.* The method is derived from the *Riemann sum of Riemann sums* formula in the last section.

We recognize the presence of functions of single variable in this function of two variables as well as the Riemann sums and the integrals of these functions in its Riemann sum and the integral.

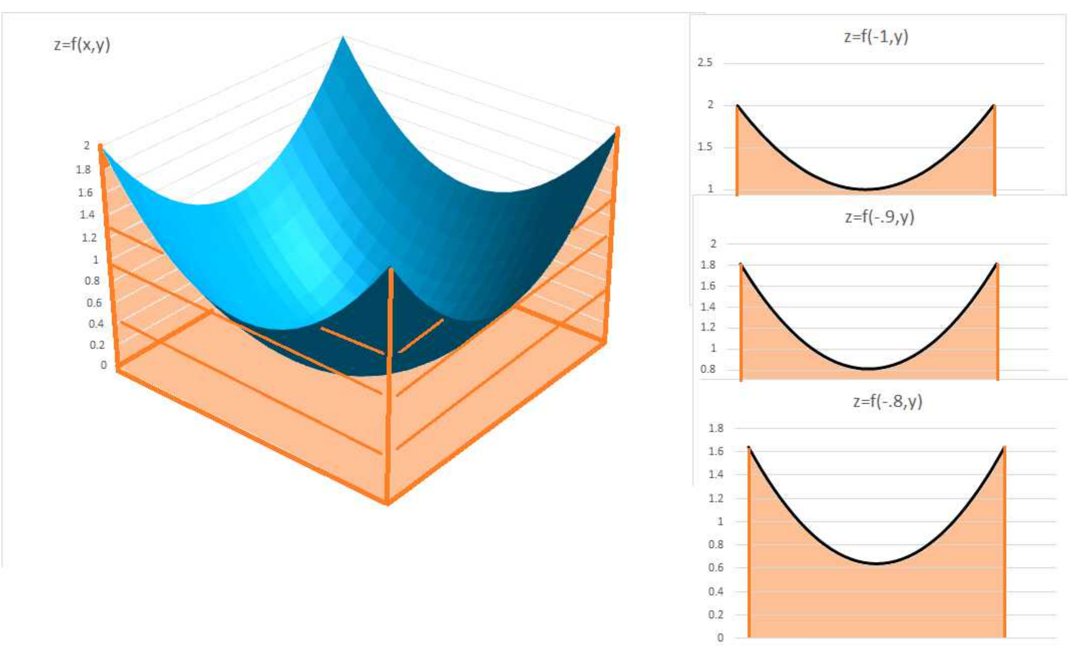
Fixing any value of x creates a new single variable function of y from $z = f(x, y)$:

$$f_x(y) = f(x, y)\,.$$

We pick one and form the Riemann integral of this function:

$$\int_c^d f_x\,dy = \int_{y=c}^{y=d} f(x, y)\,dy\,.$$

This integral is with respect to y with the domain of integration $[c, d]$. It represents the *area of the cross section* of our solid!



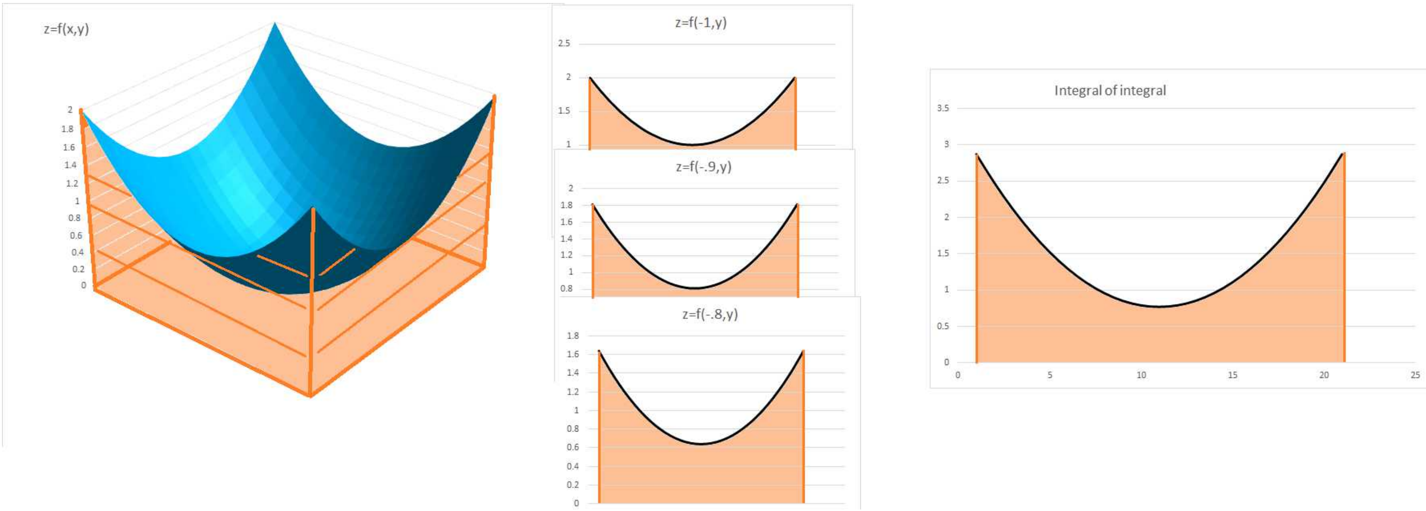
When this is done for each x, the result is a new function g defined on the interval $[a, b]$:

$$g(x) = \int_c^d f_x\,dy\,.$$

We now compute its Riemann integral:

$$\int_R f\,dxdy = \int_a^b g\,dx\,,$$

This integral is with respect to x with the domain of integration $[a, b]$. It produces the Riemann integral of the original function $z = f(x, y)$.



Alternatively, for each y, the Riemann integral is computed and, together, these numbers form a function of y. Its Riemann integral is now computed. The results are the same.

Exercise 5.3.14

Carry out the missing part of the above analysis starting with: "Fixing any value of y creates a new single variable function of x from $z = f(x, y)$...".

Our analysis amounts to the following formula similar to the formula for the second, mixed partial derivative:

Theorem 5.3.15: Fubini's Theorem: Integral of Integral

If a function $z = f(x, y)$ is integrable over a rectangle $R = [a, b] \times [c, d]$, then it is also integrable with respect to x over the interval $[a, b]$ and with respect to y over the interval $[c, d]$; moreover, we have:

$$\boxed{\begin{aligned}\iint_R f\,dA &= \int_{x=a}^{x=b} \left(\int_{y=c}^{y=d} f(x, y)\,dy \right) dx \\ &= \int_{y=c}^{c=d} \left(\int_{x=a}^{x=b} f(x, y)\,dx \right) dy\end{aligned}}$$

The right-hand sides are called *iterated integrals*. They supply as, for the first time, with a method for computing area integrals!

The method relies on an approach similar to partial differentiation in that we initially treat y as the only variable and, then, x has to be treated as a *constant*. This coordinate-by-coordinate integration may be called "partial integration".

Example 5.3.16: FTC

Fubini's Theorem allows us to use the *Fundamental Theorem of Calculus*:

$$\begin{aligned}
\iint_{[0,1]\times[0,2]} xy\,dxdy &= \int_{x=0}^{x=1}\left(\int_{y=0}^{y=2} xy\,dy\right)dx && \text{Use the CMR.}\\
&= \int_{x=0}^{x=1} x\left(\int_{y=0}^{y=2} y\,dy\right)dx && \text{Use the FTC.}\\
&= \int_{x=0}^{x=1} x\left(\frac{y^2}{2}\bigg|_{y=0}^{y=2}\right)dx && \\
&= \int_{x=0}^{x=1} x\,(2)\,dx && y \text{ is gone!}\\
&= 2\int_{x=0}^{x=1} x\,dx && \text{Use the FTC.}\\
&= 2\frac{x^2}{2}\bigg|_{x=0}^{x=1} && \\
&= 1\,. &&
\end{aligned}$$

Using "$x =,\ y =$" is optional and so are the parentheses:

$$\iint_{[0,1]\times[0,2]} xy\,dxdy = \int_0^1\int_0^2 xy\,dydx\,.$$

Non-rectangular domains of integration, such as the disk, are still to come.

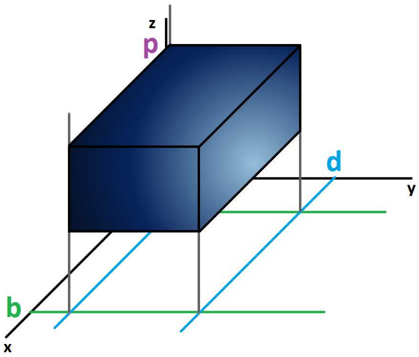
5.4. *The weight as the 3d Riemann sum*

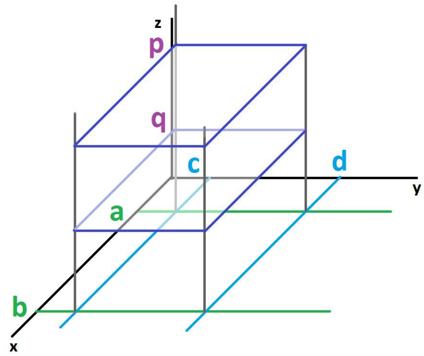
We can find volumes of various objects and, therefore, their weights – as long as the density is constant! What if the density is variable? We don't have a solution. Even simpler, what is the mass of a rectangular contained (box) filled with a gas of variables density?

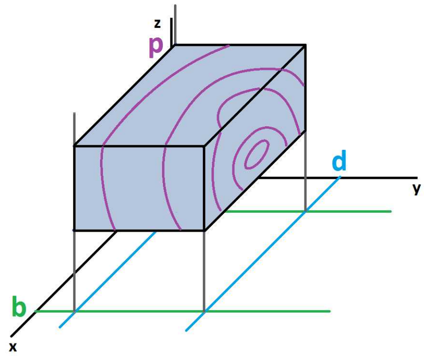
Recall the two main metaphors (in addition to that of motion) we have used for integrals:

variables	domain	metaphor #1	metaphor #2
1	interval $[a,b]$	area under the graph in $\mathbf{R}^2$	linear density
2	rectangle $[a,b]\times[c,d]$	volume under the graph in $\mathbf{R}^3$	planar density
3	box $[a,b]\times[c,d]\times[p,q]$	volume? under? the graph in $\mathbf{R}^4$?	density

It is hard to provide a similar interpretation for the 3 variables. As we have run out of dimensions to visualize these functions, we choose the second metaphor: a function of *three variables* $u = f(x,y,z)$ is understood as the density of the medium at location (x,y,z). The level surfaces below show where the density is the same.



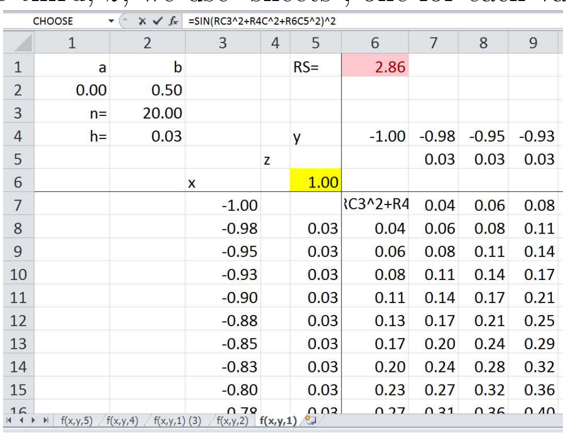




In addition to the common meaning of density, i.e., the distribution of weight, we can speak of the density of a particular material within the medium, the density of population, the temperature (the density of heat), etc.

Example 5.4.1: spreadsheet

Spreadsheets are tables of numbers and may only be good for representing functions of up to two variables, x and y. To add the third, z, we use "sheets", one for each value of z:



CHOOSE =SIN(RC3^2+R4C^2+R6C5^2)^2

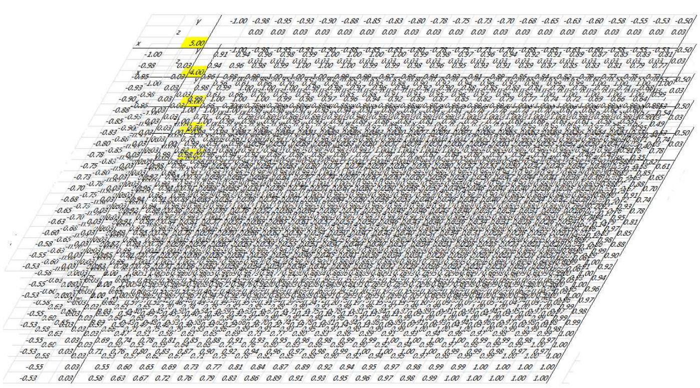
	1	2	3	4	5	6	7	8	9
1	a	b			RS=	2.86			
2	0.00	0.50							
3	n=	20.00							
4	h=	0.03			y	-1.00	-0.98	-0.95	-0.93
5				z			0.03	0.03	0.03
6			x		1.00				
7			-1.00			RC3^2+R4	0.04	0.06	0.08
8			-0.98		0.03	0.04	0.06	0.08	0.11
9			-0.95		0.03	0.06	0.08	0.11	0.14
10			-0.93		0.03	0.08	0.11	0.14	0.17
11			-0.90		0.03	0.11	0.14	0.17	0.21
12			-0.88		0.03	0.13	0.17	0.21	0.25
13			-0.85		0.03	0.17	0.20	0.24	0.29
14			-0.83		0.03	0.20	0.24	0.28	0.32
15			-0.80		0.03	0.23	0.27	0.32	0.36
16			-0.78		0.03	0.27	0.31	0.36	0.40

f(x,y,5) f(x,y,4) f(x,y,1) (3) f(x,y,2) f(x,y,1)

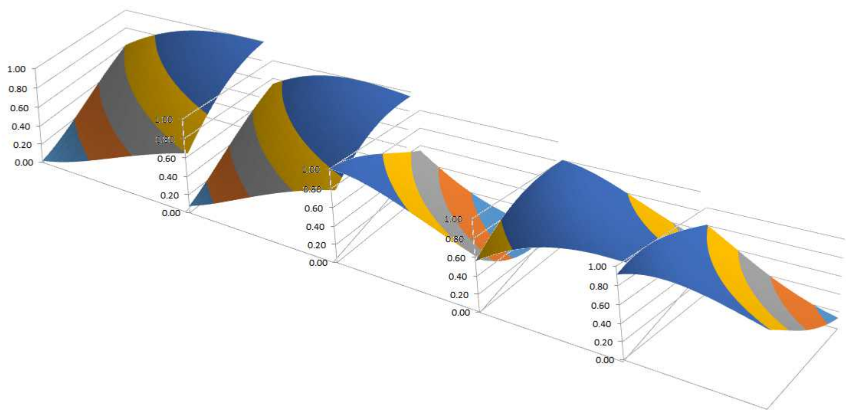
We enter the value of z at the corner of our table of values of f. The formula in each cell of this table refers – just as before – to the values of x (the first column) and y (the top row) as well as this single entry for z:

```
=SIN(RC3^2+R4C^2+R6C5^2)^2
```

A few more sheets, or layers, are created with different entries for $z = 1, 2, 3, 4, 5$. The result is a *box*, or a three-dimensional array, of numbers:



The surface shown below are the graphs of these functions of two variables: $u = f(x, y, 1), u = f(x, y, 2)$, etc.



The graph of each of these functions can again be seen to represent a *terrain*. Now, what if we think of z as *time*? Then we face a terrain changing in time! Such a process could be a wave (the sea, a drum, etc.).

The function is represented by a list of rectangular arrays. Alternatively, we see it as a rectangular array of lists...

The setup for the Riemann sums for functions of three variables is very similar to the one for two.

Suppose we have a function $u = f(x, y, z)$ defined on a box

$$B = [a, b] \times [c, d] \times [p, q], \ a < b, \ c < d, \ p < q \,.$$

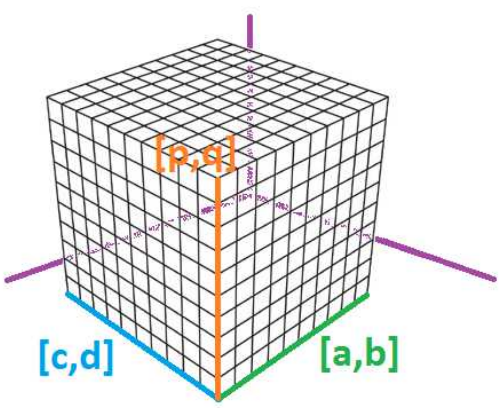
Suppose also that we have two integers $n, m, r \geq 1$. We have *three partitions* of these intervals:

axis	segment	end-points	lengths
x	$[a, b]$	$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$	$\Delta x_i = x_i - x_{i-1}$
y	$[c, d]$	$c = y_0 < y_1 < y_2 < ... < y_{m-1} < y_m = d$	$\Delta y_j = y_j - y_{j-1}$
z	$[p, q]$	$p = z_0 < z_1 < z_2 < ... < z_{n-1} < z_r = q$	$\Delta z_k = z_k - z_{k-1}$

Altogether, we have a *partition* P of the box B into smaller boxes or compartments:

$$B_{ijk} = [x_i, x_{i+1}] \times [y_j, y_{j+1}] \times [z_k, z_{k+1}] \,.$$

We know these boxes as 3-*cells*. The whole partition may look like this wire-frame:



The system is non-uniform in general.

The volume of each 3-cell is:

$$\Delta V_{ijk} = \Delta x_i \cdot \Delta y_j \cdot \Delta z_k \,.$$

In other words, the product of the increments of x, y, and z is the increment of the volume.

The points

$$X_{ijk} = (x_i, y_j, z_k), \quad i = 1, 2, ..., n, \ j = 1, 2, ..., m, \ k = 1, 2, ..., r \,,$$

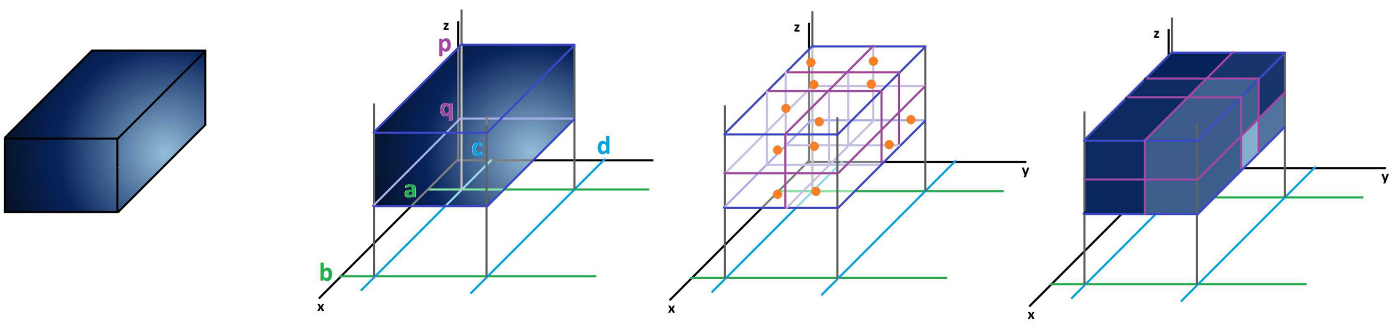
are the primary nodes, or the nodes of degree 0. What about the secondary, etc. nodes? We have a more uniform look at the nodes consistent with our view on cells. These are the 3-cells of our partition as well as the lower-dimensional cells that make up the boundaries of these cells:

cells	terms	product representation	nodes
3-cells	boxes	$[x_i, x_{i+1}] \times [y_j, y_{j+1}] \times [z_k, z_{k+1}]$	
2-cells	faces	$[x_i, x_{i+1}] \times [y_j, y_{j+1}] \times \{z_k\}$	tertiary nodes
		$[x_i, x_{i+1}] \times \{y_j\} \times [z_k, z_{k+1}]$	
		$\{x_i\} \times [y_j, y_{j+1}] \times [z_k, z_{k+1}]$	
1-cells	edges	$[x_i, x_{i+1}] \times \{y_j\} \times \{z_k\}$	secondary nodes
		$\{x_i\} \times [y_j, y_{j+1}] \times \{z_k\}$	
		$\{x_i\} \times \{y_j\} \times [z_k, z_{k+1}]$	
0-cells	vertices	$\{x_i\} \times \{y_j\} \times \{z_k\}$	primary nodes

For the weight we will only need the nodes at the 3-cells; for each triple $i = 0, 1, 2, ..., n-1$, $j = 0, 1, 2, ..., m-1$, and $k = 0, 1, 2, ..., r-1$, we have:

- a point S_{ijk} in the box B_{ijk}.

Such a combination of boxes and nodes will be called an *augmented partition* of B.



Before we address how to compute the weight, let's consider a simpler problem.

Suppose a function $y = f(X) = f(x, y, z)$ defined at S_{ijk} and gives us the *amount* of some material contained in the corresponding compartment. Then the total amount of the material in the whole box is simply the sum of the values of f.

Definition 5.4.2: sum

The *sum* of a function $z = f(x, y)$ defined at the nodes at the 3-cells of an augmented partition P of a box $R = [a, b] \times [c, d] \times [p, q]$ is defined to be the following:

$$\boxed{\sum_B f = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{r} f(S_{ijk})}$$

Note that when these nodes aren't provided, we can think of the 3-cells themselves as the inputs of the function: $S_{ijk} = B_{ijk}$. This makes f a 3-form.

Each degree 3 node gives us the density (as if constant) of the medium in the corresponding box:

$$\text{the weight of box } B_{ijk} = \underbrace{f(S_{ijk})}_{\text{density}} \cdot \overbrace{\Delta x}^{\text{depth of box}} \cdot \overbrace{\Delta y}^{\text{width of box}} \cdot \overbrace{\Delta y}^{\text{height of box}}$$

We then add all of these together in order to approximate the *weight of the solid in the box* of $u = f(x,y,z)$ inside box B. For each k, the values in the kth layer look like this:

z_k	y_0	Δy_1	y_1	Δy_2	y_2	$\dots$
x_0	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$\dots$
Δx_1	$\vert$	$f(S_{1,1,k})\Delta x_1 \Delta y_1 \Delta z_k$	$\vert$	$f(S_{1,2,k})\Delta x_1 \Delta y_2 \Delta z_k$	$\vert$	$\dots$
x_1	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$\dots$
Δx_2	$\vert$	$f(S_{2,1,k})\Delta x_2 \Delta y_1 \Delta z_k$	$\vert$	$f(S_{2,2,k})\Delta x_2 \Delta y_2 \Delta z_k$	$\vert$	$\dots$
x_2	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$\dots$
$\dots$	.	..	.	..	.	$\dots$.

Definition 5.4.3: Riemann sum

The *Riemann sum* of a function $u = f(x,y,z)$ defined at the nodes at the 3-cells of an augmented partition P of a box $B = [a,b] \times [c,d] \times [p,q]$ is defined to be

$$\boxed{\sum_B f\,\Delta x \Delta y \Delta z = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{r} f(S_{ijk}) \Delta x_i \Delta y_j \Delta z_k}$$

The abbreviated formula is:

$$\sum_B f\,\Delta x \Delta y \Delta z = \sum_{i,j,k} f(S_{ijk}) \Delta V_{ijk}$$

Just as before, we are allowed to have negative values of f. Furthermore, we can have negative lengths of the intervals for the independent variables. The box $[a,b] \times [c,d] \times [p,q]$ has a positive orientation whenever $a < b$, $c < d$, $p < q$; in other words, when all three segments have positive orientations within their respective axes. A reversal of the orientation of a single segment reverses the orientation of the box.

Exercise 5.4.4

What if the orientations of *two* segments are reversed?

5.5. *The weight as the 3d Riemann integral*

Now, if the density function varies continuously, the Riemann sums are just approximations of the actual weight and, in order to improve them, we have to refine the partition:

$$n, m, r \to \infty \quad \text{and} \quad \Delta x_i, \Delta y_j, \Delta z_k \to 0\,.$$

Just as before, we define the *mesh of a partition* P as follows:

$$|P| = \max_{i,j,k} \{\Delta x_i,\ \Delta y_j, \Delta z_k\}\,.$$

It is a measure of “refinement” of P.

Definition 5.5.1: Riemann integral

The *Riemann integral* of a function $u = f(x,y,z)$ over box $B = [a,b]\times[c,d]\times[p,q]$ is defined to be the limit of a sequence of its Riemann sums with the mesh of their augmented partitions P_s approaching 0 as $s \to \infty$. When all these limits exist and are all equal to each other, f is called an *integrable function over* B and the result is denoted as follows:

$$\iiint_B f(x,y,z)\,dxdydz = \lim_{s\to\infty} \sum_B f_s\,\Delta x \Delta y \Delta z$$

where f_s is f sampled over the partition P_s. It is also called the *definite integral*, the *triple integral*, or simply *integral*, of f over B. When all these limits are equal to $+\infty$ (or $-\infty$), we say that the integral is *infinite* and write:

$$\iiint_B f(x,y,z)\,dxdydz = +\infty \ (\text{or } \ -\infty)\,.$$

An abbreviated notation is:

Triple integral

$$\iiint_B f\,dV$$

As you can see, $dxdydz$ is replaced with dV, where "V" stands for "volume".

Theorem 5.5.2: Constant Integral Rule

Suppose $z = f(x,y,z)$ is constant on box $B = [a,b]\times[c,d]\times[p,q]$, i.e., $f(x,y,z) = m$ for all (x,y,z) in B for some real number m. Then f is integrable over B and

$$\iiint_B f\,dV = m(b-a)(d-c)(q-p)$$

Theorem 5.5.3: Cont. => Integr.

All continuous functions on box B are integrable on B.

Proof.

The converse isn't true.

The properties of the Riemann integral mimic the ones for numerical functions (and functions of two variables).

The interpretation of additivity is the same: the quantity in a region formed from two regions with a negligible overlap is equal to the sum of the quantities in the two.

It's as if we put a divider in our container...

Theorem 5.5.4: Additivity For Triple Integrals

Suppose a function $u = f(x, y, z)$ *is integrable over the boxes*

$$R = [a, q] \times [c, d] \times [p, q] \quad \text{and} \quad S = [q, b] \times [c, d] \times [p, q]\,.$$

Then f *is integrable over the box* $R \cup S = [a, b] \times [c, d] \times [p, q]$ *and we have:*

$$\boxed{\iiint_R f\, dV + \iiint_S f\, dV = \iiint_{R\cup S} f\, dV}$$

Thus, adding the domains of integration adds the integrals too.

Theorem 5.5.5: Local Integrability For Triple Integrals

If $u = f(x, y, z)$ *is integrable over box* B *then it is also integrable over any box* B' *in* B.

Theorem 5.5.6: Cont. => Integr.

All piecewise continuous functions are integrable.

Just as before, the box with a larger density weighs more:

Theorem 5.5.7: Comparison Rule For Triple Integrals

If

$$f(x, y, z) \geq f(x, y, z) \quad \text{for all } (x, y, z) \text{ in box } B\,,$$

then

$$\boxed{\iiint_B f\, dV \geq \iiint_B f\, dV}$$

provided $z = f(x, y, z)$ *and* $z = g(x, y, z)$ *are integrable functions over* B. *Otherwise we have:*

$$\boxed{\begin{array}{lcl} \iiint_B f\, dV = -\infty & \Longrightarrow & \iiint_B g\, dV = -\infty \\ \iiint_B f\, dV = +\infty & \Longleftarrow & \iiint_B g\, dV = +\infty \end{array}}$$

If we know only estimates of the density, we have an estimate – below and above – for the weight of the box.

Theorem 5.5.8: Estimate Rule For Triple Integrals

Suppose $u = f(x, y, z)$ *is an integrable function over box* $B = [a, b] \times [c, d] \times [p, q]$. *Then, if* $a < b$, $c < d$, $p < q$, *and*

$$m \leq f(x, y, z) \leq M\,,$$

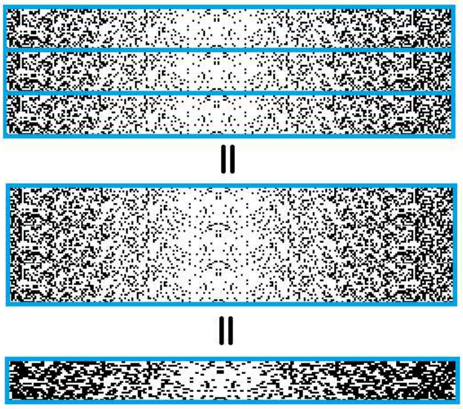
for all (x,y,z) in B, we have:

$$m(b-a)(d-c)(q-p) \leq \iiint_R f\, dV \leq M(b-a)(d-c)(q-p)$$

Exercise 5.5.9

What if the orientation of the box is negative?

Finally, these are the algebraic properties. First, the picture below illustrates the idea that tripling the density of a solid will triple its weight:

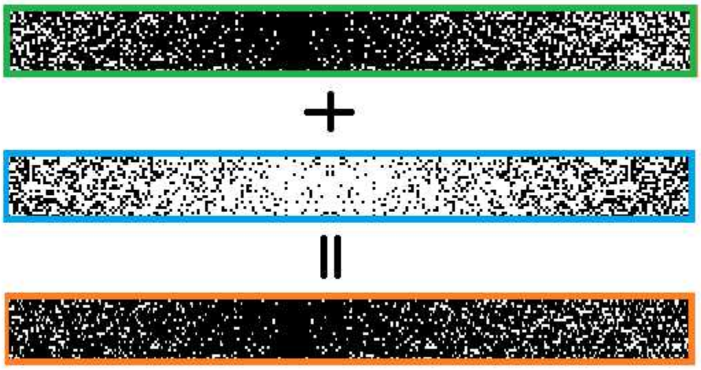


Theorem 5.5.10: Constant Multiple Rule For Triple Integrals

Suppose $u = f(x,y,z)$ is an integrable function over a box B. Then so is $c \cdot f$ for any real c and we have:

$$\iiint_B (c \cdot f)\, dV = c \cdot \iiint_B f\, dV$$

Second, the picture below illustrates that when gas is pumped from one container to another that already contains gas, the resulting density is the sum of the two:



Theorem 5.5.11: Sum Rule For Triple Integrals

Suppose $u = f(x,y,z)$ and $u = g(x,y,z)$ are integrable functions over a box B.

Then so is $f+g$ and we have:

$$\boxed{\iiint_B (f+g)\, dV = \iiint_B f\, dV + \iiint_B g\, dV}$$

Exercise 5.5.12

Prove these theorems.

The actual computations of integrals require *single integrals*. We recognize the presence of functions of single variable in this function of three variables as well as the Riemann sums and the integrals of these functions in its Riemann sum and the integral.

Theorem 5.5.13: Fubini's Theorem: Integral of Integral of Integral

If a function $u = f(x,y,z)$ is integrable over a box $B = [a,b] \times [c,d] \times [p,q]$, then it is also integrable with respect to x over the interval $[a,b]$, with respect to y over the interval $[c,d]$, and with respect to z over the interval $[p,q]$; moreover, we have (in any order of variables):

$$\boxed{\iiint_B f\, dV = \int_{x=a}^{x=b} \left(\int_{y=c}^{y=d} \left(\int_{z=p}^{z=q} f(x,y,z)\, dz \right) dy \right) dx}$$

Example 5.5.14: a computation

We gradually progress from inside out:

$$\begin{aligned}
\iiint_{[0,1]\times[0,2]\times[0,3]} x^3y^2z\, dV &= \int_{x=0}^{x=1} \left(\int_{y=0}^{y=2} \left(\int_{z=0}^{z=3} x^3y^2z\, dz \right) dy \right) dx \\
&= \int_{x=0}^{x=1} \left(\int_{y=0}^{y=2} x^3y^2 \left(\int_{z=0}^{z=3} z\, dz \right) dy \right) dx \\
&= \int_{x=0}^{x=1} \left(\int_{y=0}^{y=2} x^3y^2 \left(\frac{z^2}{2}\bigg|_{z=0}^{z=3} \right) dy \right) dx \\
&= \int_{x=0}^{x=1} \left(\int_{y=0}^{y=2} x^3y^2 \left(\frac{9}{2} \right) dy \right) dx && \text{z is gone!} \\
&= \int_{x=0}^{x=1} \frac{9}{2}x^3 \left(\int_{y=0}^{y=2} y^2\, dy \right) dx \\
&= \int_{x=0}^{x=1} \frac{9}{2}x^3 \left(\frac{y^3}{3}\bigg|_{y=0}^{y=2} \right) dx \\
&= \int_{x=0}^{x=1} \frac{9}{2}x^3 \left(\frac{8}{3} \right) dx && \text{y is gone!} \\
&= 12 \int_{x=0}^{x=1} x^3\, dx \\
&= 12 \frac{x^4}{4}\bigg|_{x=0}^{x=1} \\
&= 12 \frac{1}{4} && \text{x is gone too!} \\
&= 3\,.
\end{aligned}$$

Specifying the variables of the bounds of integration as well as using the parentheses is optional.

Example 5.5.15: another computation

Alternative notation:

$$\begin{aligned}
\iiint_{[0,1]\times[0,2]\times[0,3]} x^3y^2z\,dV &= \int_0^1 \int_0^2 \int_0^3 x^3y^2z\,dz \quad dy \quad dx\\
&= \int_0^1 \int_0^2 x^3y^2 \int_0^3 z\,dz \quad dy \quad dx\\
&= \int_0^1 \int_0^2 x^3y^2 \left.\frac{z^2}{2}\right|_0^3 \quad dy \quad dx\\
&= \int_0^1 \int_0^2 x^3y^2 \frac{9}{2} \quad dy \quad dx\\
&= \int_0^1 \frac{9}{2}x^3 \int_0^2 y^2 \quad dy \quad dx\\
&= \int_0^1 \frac{9}{2}x^3 \left.\frac{y^3}{3}\right|_0^2 \quad dx\\
&= \int_0^1 \frac{9}{2}x^3 \frac{8}{3} \quad dx\\
&= 12\int_0^1 x^3\,dx\\
&= 12\left.\frac{x^4}{4}\right|_0^1\\
&= 12\frac{1}{4}\\
&= 3\,.
\end{aligned}$$

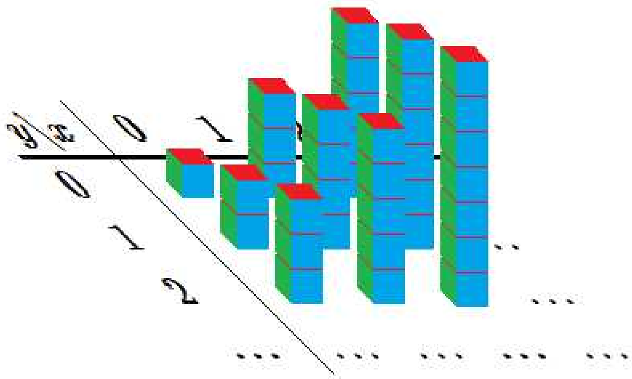
Non-rectangular domains, such as the ball, are still to come.

5.6. Lengths, areas, volumes, and beyond

What do the types of Riemann sums and integrals that we have considered have in common?

Example 5.6.1: average height

Suppose we would like to find the average height of a building in a city. For simplicity, we assume that there is one building in each block. The city might look like this:



For our computations we collect the heights of these building and put them in a table, which makes a *function of two variables* f. These may be its inputs and outputs:

$y\backslash x$	0	1	2	...	10
0	$(0,0)$	$(0,1)$	$(0,2)$	...	$(0,10)$
1	$(1,0)$	$(1,1)$	$(2,2)$	...	$(1,10)$
2	$(2,0)$	$(1,1)$	$(2,2)$	...	$(2,10)$
...	...	...	...	...	
10	$(10,0)$	$(10,1)$	$(10,2)$	...	$(10,10)$

leading to

$y\backslash x$	0	1	2	...	10
0	1	3	5	...	0
1	2	4	6	...	0
2	3	5	7	...	1
...	...	...	...	...	
10	0	1	7	...	0

This may look like a generic function of two variables $z = f(x, y)$, but let's take a closer look at how the data. It's not a single number for each location but for each block, i.e., a *rectangle*. The data then is represented by a table:

$y\backslash x$	0		1		2		3	...	9		10
0	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$
	$\vert$	1	$\vert$	3	$\vert$	5	$\vert$	...	$\vert$	0	$\vert$
1	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$
	$\vert$	2	$\vert$	4	$\vert$	6	$\vert$	...	$\vert$	0	$\vert$
2	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$
	$\vert$	3	$\vert$	5	$\vert$	7	$\vert$	...	$\vert$	1	$\vert$
3	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$
...		...		...		...		...		...	
9	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$
	$\vert$	0	$\vert$	1	$\vert$	7	$\vert$	...	$\vert$	0	$\vert$
10	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	$--$	$\bullet$	...	$\bullet$	$--$	$\bullet$

We realize that this isn't just a function of two variables; it's a *discrete* 2-*form*! Then to find the average height, we find the total of the heights of the buildings, which is the *sum* of f over the rectangle $R = [0, 10] \times [0, 10]$:

$$\text{total height } = \sum_R f\,,$$

and then divide by the number of buildings:

$$\text{average height } = \frac{1}{10 \cdot 10} \sum_R f\,.$$

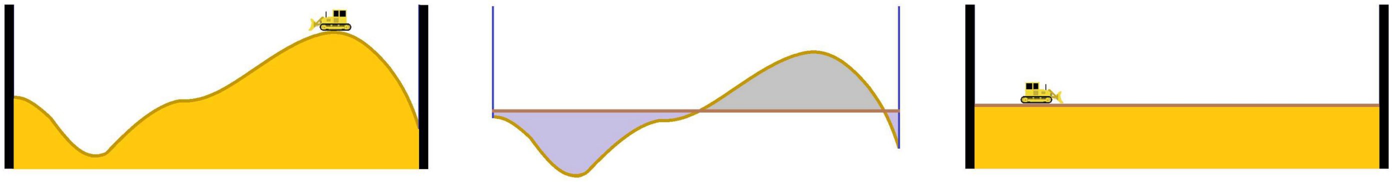
However, a more complex situation is that of a city with blocks of various dimensions. In that case, the average height is the total of the *volumes* of the buildings, which is the Riemann sum of the heights over R:

$$\text{total volume } = \sum_{ij} f(X_{ij})\,\Delta A_{ij} = \sum_R f\,\Delta x \Delta y\,,$$

divided by the area of the city, 100:

$$\text{average height } = \frac{1}{\text{area of } R}\sum_R f\,\Delta x \Delta y\,.$$

The analysis applies to the idea of the average amount of any *material spread over a region.*



Some are familiar for the three cases of dimensions $1,2,3$:

$$\begin{aligned}
\text{average density of segment } I &= \frac{1}{\text{length of } I}\sum_I f\,\Delta x\\
\text{average density of rectangle } R &= \frac{1}{\text{area of } R}\sum_R f\,\Delta x \Delta y\\
\text{average density of box } B &= \frac{1}{\text{volume of } B}\sum_B f\,\Delta x \Delta y\,\Delta z
\end{aligned}$$

Here, the meaning varies:

dim	domain	space	measure	f is...	$\sum_U f\,\Delta x...$ is...
1	interval	in $\mathbf{R}^1$	length	linear density	total amount
2	rectangle	in $\mathbf{R}^2$	area	planar density	total amount
3	box	in $\mathbf{R}^3$	volume	density	total amount
...					

Do we continue?

To proceed to the n-dimensional case, we first need to understand how to *measure* the size of this n-dimensional "box". Recall that an n-cell in $\mathbf{R}^n$ is the set of all points each coordinate of which lies within a predetermined interval of values. In other words, for any list of n pairs of real numbers,

$$a_1 < b_1,\ a_2 < b_2,\ ...,\ a_n < b_n\,,$$

we define the box to be

$$B = \{X = (x_1, x_2, ..., x_n) :\ a_i \le x_i \le b_i,\ i = 1, 2, ..., n\}\,.$$

The notation we have used is:

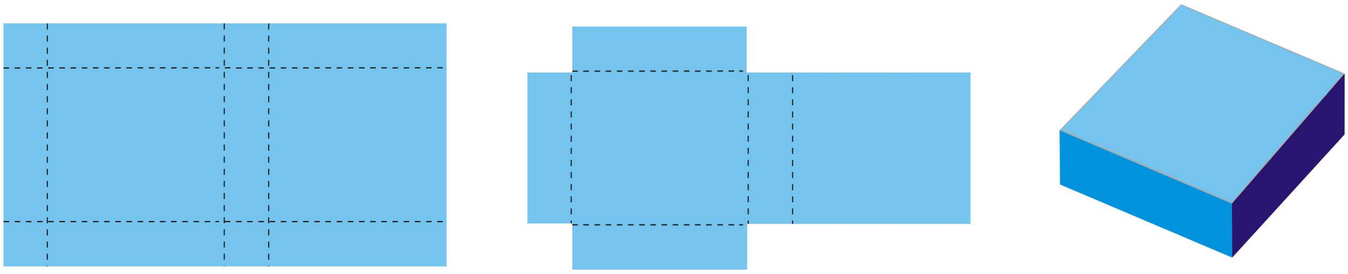
Box

$$B = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$$

The *n-volume* of the box B is simply the product of the lengths of the intervals that make it up:

$$v(B) = (b_1 - a_1)(b_2 - a_2)...(b_n - a_n)\,.$$

Now, the *faces* of this n-box are $(n-1)$-cells.



Together they form the *boundary* of the n-cell.

It is important that the n-volume of an $(n-1)$-cell is zero! That is why we can create a partition of an n-cell into smaller n-cell as long as they only intersect by their faces.

This is the nth row in the table above:

dim	domain	space	measure	f is...	$\sum_B f\,\Delta V$ is...
n	cell	in $\mathbf{R}^n$	n-volume	density	total amount

Now, what is the meaning of the Riemann sum $\sum_B f\,\Delta V$?

The n-cell B is the product of the intervals located in the axes of $\mathbf{R}^n$. We provide a partition for each. These partitions cut B into smaller n-cells creating a *partition* P. A discrete n-form f is then a function that assigns a number to each of these cells.

Definition 5.6.2: Riemann sum

The *Riemann sum* of a discrete n-form f over a partition of an n-cell B is defined to be the following:

$$\boxed{\sum_B f\,\Delta V = \sum_C f(C)v(C)}$$

where summation is over all n-cells C of the partition.

Definition 5.6.3: average value of form

The *average value of an n-form f* defined over a partition of an n-cell B is defined to be the following:

$$\boxed{\text{average value of } f \ = \frac{1}{v(B)}\sum_B f\,\Delta V}$$

The limit of the Riemann sums, as the mesh of the partition is approaching 0, is the Riemann integral. For

all three dimensions, the integral of f is the total weight:

$$\begin{aligned} \text{weight} &= \int_{[a,b]} f(x)\,dx \\ \text{weight} &= \iint_{[a,b]\times[c,d]} f(x,y)\,dxdy \\ \text{weight} &= \iiint_{[a,b]\times[c,d]\times[p,q]} f(x,y,z)\,dxdydz \end{aligned}$$

Furthermore, the *average density* is the total amount (weight) divided by the measure of the "container":

$$\begin{aligned} \text{average density} &= \frac{1}{b-a}\int_{[a,b]} f(x)\,dx \\ \text{average density} &= \frac{1}{(b-a)(d-c)}\iint_{[a,b]\times[c,d]} f(x,y)\,dxdy \\ \text{average density} &= \frac{1}{(b-a)(d-c)(q-p)}\iiint_{[a,b]\times[c,d]\times[p,q]} f(x,y,z)\,dxdyd. \end{aligned}$$

This is a launching pad into higher dimensions! We can now understand these three as one.

Definition 5.6.4: average value of function

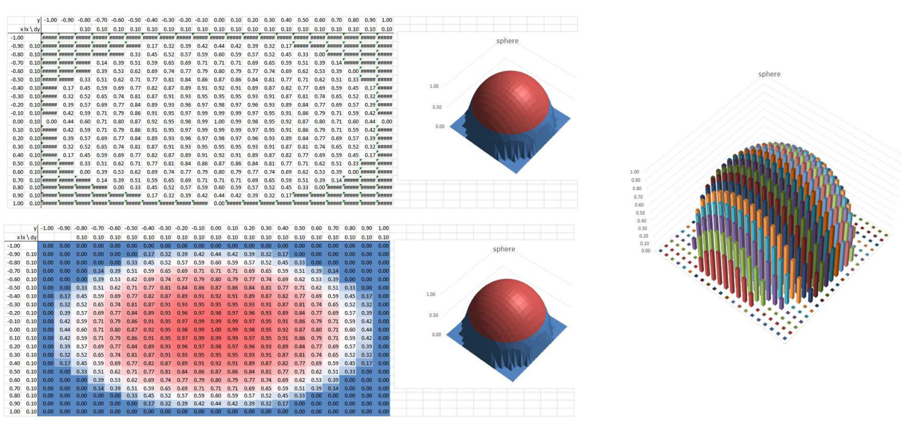
The *average value of a function* $z = f(X)$ of n variables over an n-cell B is defined to be the following:

$$\boxed{\text{average value of } f \;=\; \frac{1}{n\text{-volume of } B}\int_B f(X)\,dV}$$

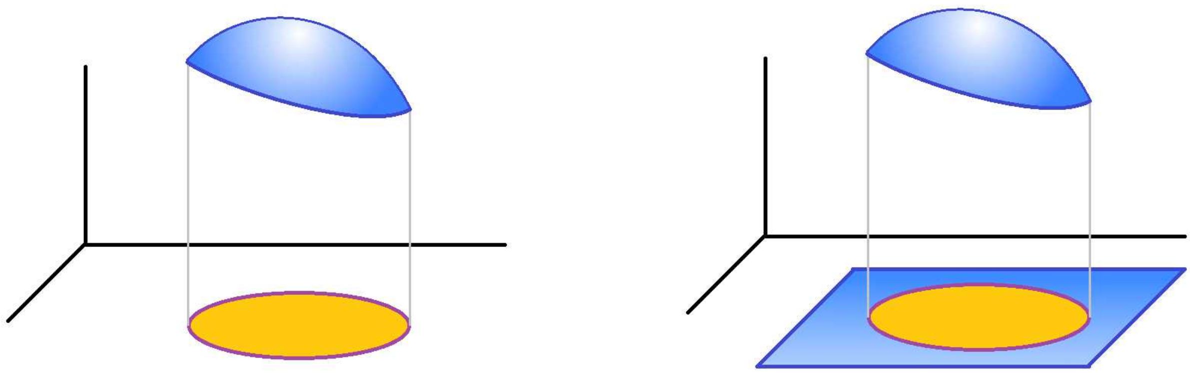
5.7. Outside the sandbox

So far, the footprints of our solids are just rectangles. The example of the sphere in the beginning of the chapter shows that yet another generalization of the concept is needed: we need to be able to handle non-rectangular regions too.

Recall how we computed the volume of the sphere:



The example also suggest the way to deal with this: we fill the missing values with 0s.



Definition 5.7.1: Riemann integral

Suppose a function $z = f(x, y)$ is defined over an arbitrary region U in $\mathbf{R}^2$. Then the *Riemann integral* of f over U is defined to be the integral over any rectangle R that contains U,

$$\boxed{\iint_U f(x,y)\,dxdy = \iint_R g(x,y)\,dxdy}$$

with the function g defined by:

$$g(x,y) = \begin{cases} f(x,y) & \text{inside } B\,, \\ 0 & \text{outside } U\,. \end{cases}$$

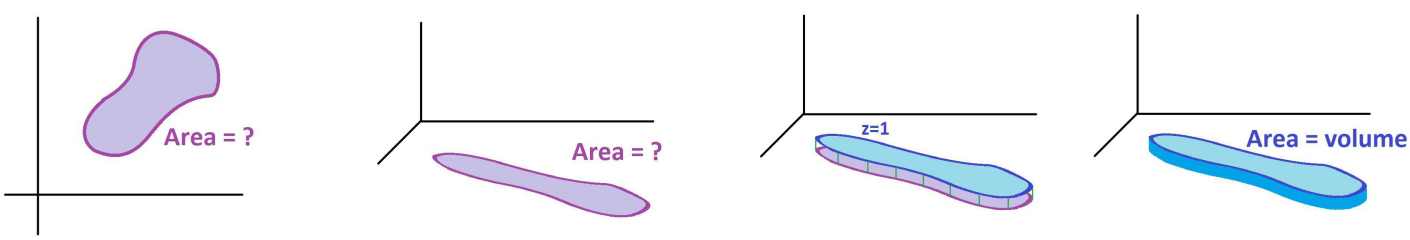
When such integral exists, the function f is called *integrable over* U.

Of course, this integral, and the area, might be infinite or not exist at all...

Furthermore, the ability to compute the *volumes* of complex solids presupposes the ability to compute the *areas* of complex plane regions!

But what is area anyway? The answer has been "the area under the graph of a function of one variable". The answer leaves out a lot of regions such a disk. Fortunately, with the above definition we can have a double integral over any region U. In the simplest case we pick f to be constant at 1 over U. Then

$$\text{the volume of } U \;=\; \text{the area of } U \;\cdot\; \text{the thickness}\,.$$



Definition 5.7.2: area of region

Suppose U is a region in the plane. Then the *area* $A(U)$ of U is defined to be the Riemann integral of the function equal to 1 over U:

$$\boxed{A(U) = \iint_U 1\,dxdy}$$

Every integral is a limit and a limit might not exist; therefore, some regions have no areas!

Warning!

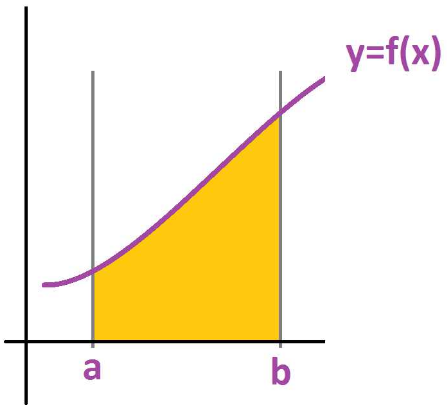
It's not the same as to say that some regions have *zero* areas.

Note that the definition settles the debt from Volume 3 (Chapter 3IC-3) about the meaning of the volume of a "shell".

Thus we can finally put to rest the early idea we have relied on for so long that the area is a single integral; *the area is a double integral*! It's only when the region has a *special* shape, this double integral turns into a single one.

Indeed, let's match the new and the old definitions. Suppose $y = f(x)$ is a function. Then the area U located under its graph between the lines $x = a$ and $x = b$ is equal to the following:

$$\int_a^b f\,dx \text{ but also } \iint_U 1\,dxdy\,.$$



How do we represent this region U algebraically? We know that x runs between its *bounds* a and b, what about y? Does it simply run between some c and d? No, that would make U a rectangle!

The answer is: the bounds of y, in the pair (x, y), depend on x.

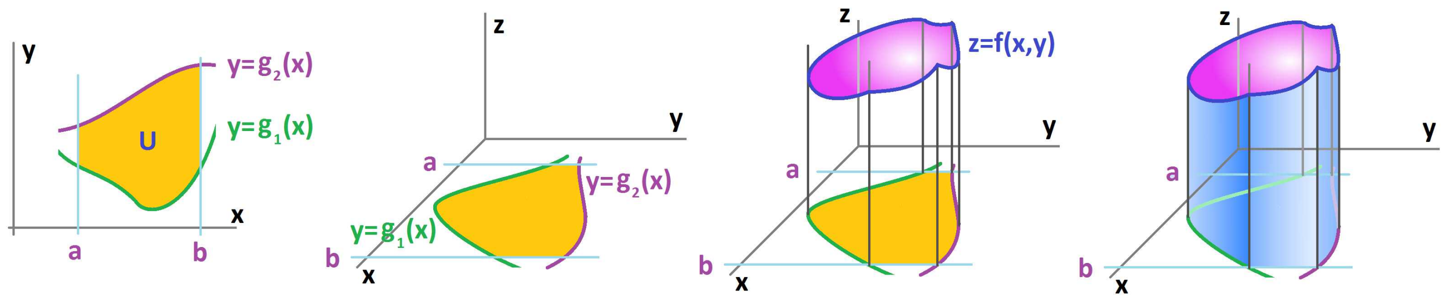
In other words, we fix x and look at all possible values of y: they lie between 0 and $f(x)$. So, we have:

$$U = \{(x, y) : a \le x \le b,\ 0 \le y \le f(x)\}\,.$$

These four will serve as the bounds in our double integral in the new version of Fubini's Theorem. Let's carry out the computation to confirm the match:

$$\begin{aligned}
\iint_U 1\,dxdy &= \int_{x=a}^{x=b} \left(\int_{y=0}^{y=f(x)} 1\,dy \right) dx \\
&= \int_{x=a}^{x=b} \left(y \Big|_{y=0}^{y=f(x)} \right) dx \\
&= \int_{x=a}^{x=b} (f(x) - 0)\ dx \\
&= \int_a^b f(x)\,dx, \quad \text{indeed!}
\end{aligned}$$

The case we will address is only slightly more complex.



Theorem 5.7.3: Fubini's Theorem For Two Variables

Suppose $z = f(x,y)$ is a function integrable of the plane region given by

$$\begin{aligned} U &= \{(x,y) : & a \le x \le b, \\ & & g_1(x) \le y \le g_2(x)\} . \end{aligned}$$

Then

$$\boxed{\iint_U f\, dxdy = \int_{x=a}^{x=b} \left(\int_{y=g_1(x)}^{y=g_2(x)} f(x,y)\, dy \right) dx}$$

Just as before, the integral in the parentheses is the area $A(x)$ of the cross-section of the solid that lies above the region:

$$\iint_U f\, dxdy = \int_{x=a}^{x=b} A(x)\, dx .$$

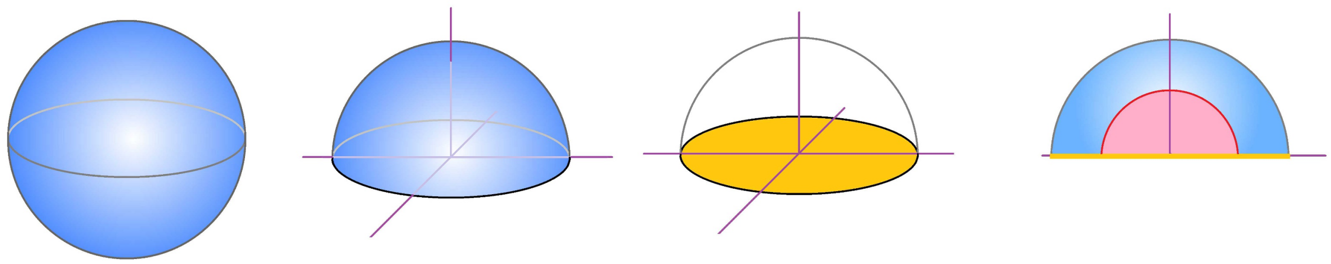
Therefore, the theorem is just a special case of the *Cavalieri Principle*.

So, to set up such an integral we should choose which of the two variables: x is the *independent* variable and y is the dependent. As an independent variable, x can vary freely between two fixed numbers while the bounds for y might depend on x.

Example 5.7.4: sphere

Let's go back to the beginning of the chapter and compute, exactly this time, the volume of the sphere of radius R. It is given by the implicit relation $x^2 + y^2 + z^2 = R^2$. We represent the upper half of the sphere as the region bounded by the xy-plane from below and by the graph of the function from above:

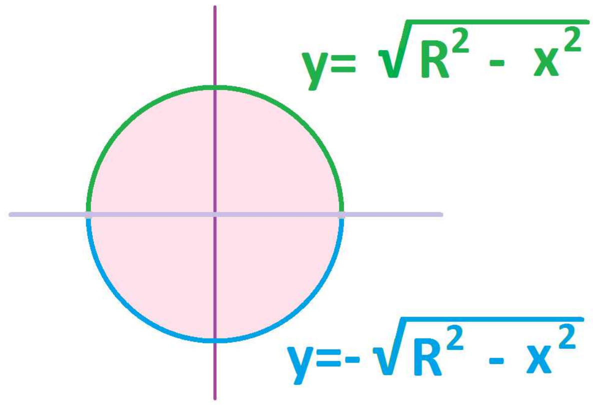
$$f(x,y) = \sqrt{R^2 - x^2 - y^2} .$$



Its domain is the "footprint", or the shadow, of the surface:

$$U = \{(x,y) : x^2 + y^2 \le R^2\} .$$

It will also serve as the domain of integration!



But in order to apply *Fubini's Theorem* we need to represent it in the appropriate way, with explicit bounds for x and y. First we include all possible values of x and then, for a fixed x, we find the interval for y (visible on the picture):

$$\begin{aligned} U &= \{(x,y): \ -R \le x \le R, \\ &\qquad\qquad -\sqrt{R^2-x^2} \le y \le \sqrt{R^2-x^2}\}\,. \end{aligned}$$

Now the volume is this integral:

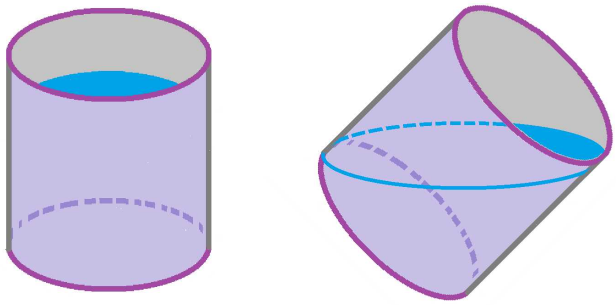
$$\begin{aligned}
\text{Volume} &= \iint_U \sqrt{R^2-x^2-y^2}\,dA && \text{Fubini.}\\
&= \int_{x=-R}^{x=R} \left(\int_{y=-\sqrt{R^2-x^2}}^{y=\sqrt{R^2-x^2}} \sqrt{R^2-x^2-y^2}\,dy \right) dx && \text{A familiar integral.}\\
&= \int_{x=-R}^{x=R} \left(\frac{y}{2}\sqrt{R^2-x^2-y^2} + \frac{R^2-x^2}{2}\sin^{-1}\frac{y}{R^2-x^2} \Bigg|_{y=-\sqrt{R^2-x^2}}^{y=\sqrt{R^2-x^2}} \right) dx\\
&= \int_{x=-R}^{x=R} \frac{R^2-x^2}{2}(\pi/2-(-\pi/2))\,dx\\
&= \pi/2 \int_{x=-R}^{x=R} (R^2-x^2)\,dx\\
&= \pi/2(R^2x - x^3/3)\Bigg|_{x=-R}^{x=R}\\
&= \pi(R^2R - R^3/3)\\
&= \frac{2}{3}\pi R^3\,.
\end{aligned}$$

The integral of y in parentheses is for a fixed $x = c$ and represents the area of the corresponding cross-section of the half-sphere with this vertical plane.

Thus the complexity of solving such a problem lies more with finding the bounds of integration than the function to integrate. However, the task is familiar from Volume 3 (Chapter 3IC-3).

Exercise 5.7.5

How much water is in a bucket that is tilted so the degree that makes the water to spill out? Solve (a) without integration, and (b) with integration.

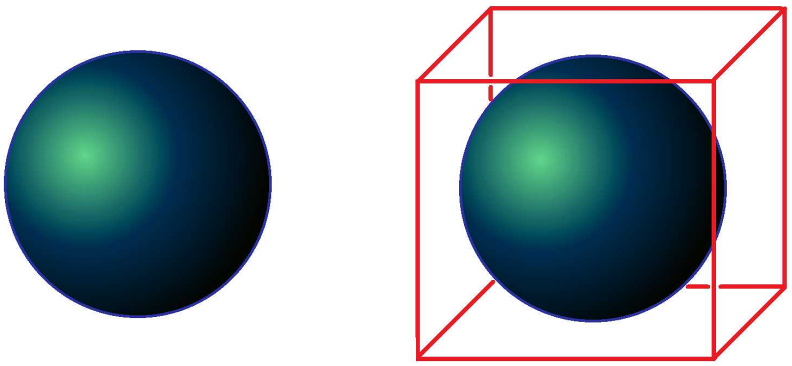


Exercise 5.7.6

What geometric properties of the bucket allow one to solve the problem without integration?

5.8. Triple integrals

The idea of integration over 3d solids mode complex than boxes is similar to that for 3d integrals. We put our solid inside one:



Definition 5.8.1: Riemann integral

Suppose a function $z = f(X)$ is defined over an arbitrary region U in $\mathbf{R}^n$. Then the *Riemann integral* of f over U is defined to be the integral over any box B that contains U,

$$\boxed{\int_U f(X)\, dV = \int_B g(X)\, dV}$$

with the function g defined by:

$$g(X) = \begin{cases} f(X) & \text{inside } B\,, \\ 0 & \text{outside } U\,. \end{cases}$$

When such integral exists, the function f is called *integrable over* U.

The following makes sense now:

Definition 5.8.2: n-volume

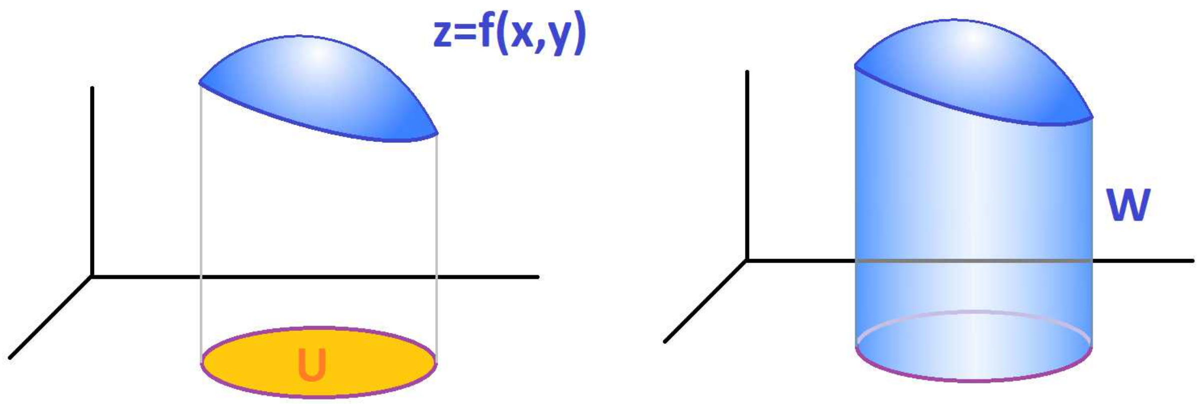
Suppose U is a region in $\mathbf{R}^n$. Then the *n-volume* $V(U)$ of U is defined to be the Riemann integral of the function equal to 1 over U:

$$\boxed{V(U) = \int_U 1\, dV}$$

Thus we can put to rest the recent idea that the volume is a double integral; *the volume is a triple integral*! It's only when the region has a *special* shape, this triple integral turns into a double one.

Indeed, let's match the new and the old definitions. Suppose $z = f(x, y)$ is a function. Then the 3d region W located under its graph above a plane region U is:

$$\iint_U f\, dxdy \quad \text{but also} \quad \iiint_W 1\, dxdydz\,.$$



How do we represent this region W algebraically? We know that (x, y) runs within W, what about z? Does it run between some p and q? No, that would make W a cylinder!

The answer is: The bounds of z, in the triple (x, y, z), depend on (x, y).

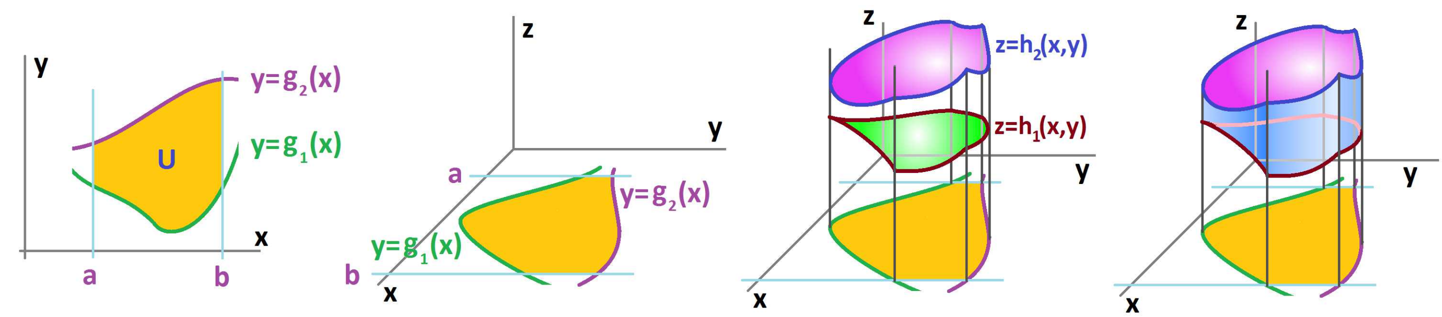
In other words, we fix (x, y) in U and look at all possible values of z: they lie between 0 and $f(x, y)$. So, we have:

$$W = \{(x, y, z) : (x, y) \text{ in } U,\ 0 \le z \le f(x, y)\}\,.$$

Let's carry out the computation to confirm the match:

$$\begin{aligned}
\iiint_W 1\, dxdydz &= \int_U \left(\int_{z=0}^{y=f(x,y)} 1\, dz \right) dxdy \\
&= \int_U \left(z \Big|_{z=0}^{z=f(x,y)} \right) dxdy \\
&= \int_U \left(f(x, y) - 0 \right) dxdy \\
&= \int_U f(x, y)\, dxdy, \quad \text{indeed!}
\end{aligned}$$

The case we will address is only slightly more complex: instead of one surface bounding from above we have two bounding from above and below.



We accept the following without proof:

Theorem 5.8.3: Incremental Fubini's Theorem

Suppose $u = f(x, y, z)$ *is a function integrable of the 3d region given by*

$$W = \{(x, y, z) : (x, y) \text{ in } U,\ h_1(x, y) \leq z \leq h_2(x, y)\},$$

where U *is some plane region. Then*

$$\boxed{\iiint_W f\, dxdydz = \int_U \left(\int_{z=h_1(x,y)}^{z=h_2(x,y)} f(x, y, z)\, dz \right) dxdy}$$

Let's make the region U in the xy-plane specific. It is in fact identical to the one we considered above: bounded between two graphs:

$$U = \{(x, y) : a \leq x \leq b,\ g_1(x) \leq y \leq g_2(x)\}.$$

In other words, we take the 2-*dimensional Fubini's Theorem* and add another integral in both left- and right-hand sides. The new integral – with respect to z – will need its own bounds...

Theorem 5.8.4: Fubini's Theorem For Three Variables

Suppose $u = f(x, y, z)$ *is a function integrable of the 3d region given by*

$$\begin{aligned} W \ = \{(x, y, z) : \ & a \leq x \leq b, \\ & g_1(x) \leq y \leq g_2(x), \\ & h_1(x, y) \leq z \leq h_2(x, y)\}. \end{aligned}$$

Then

$$\boxed{\iiint_W f\, dxdy = \int_{x=a}^{x=b} \left(\int_{y=g_1(x)}^{y=g_2(x)} \left(\int_{z=h_1(x,y)}^{z=h_2(x,y)} f(x, y)\, dz \right) dy \right) dx}$$

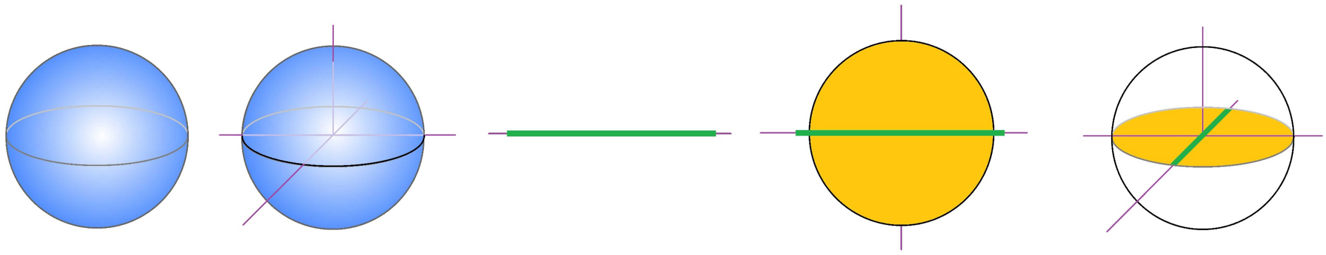
So, to set up such an integral we should choose the order of the three variables: x is the *independent* variable, y is dependent on x, and z is dependent on x *and* y. As an independent variable, x can vary freely between two fixed numbers while the bounds for y might depend on x and the bounds for z on both x *and* y. The order of variables may change though...

Example 5.8.5: sphere

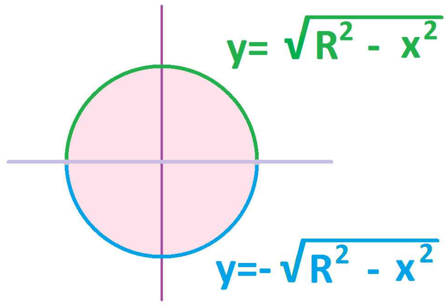
Let's take another look at the sphere. The computation of its volume in the last section – as a double integral – *transforms* into a new one – as a triple integral. We consider the *whole* sphere this time (we don't have to rely on symmetry anymore!). We know that we represent the lower half and the upper

half of the surface of the sphere as follows:

$$h_1(x,y) = -\sqrt{R^2 - x^2 - y^2} \quad \text{and} \quad h_2(x,y) = \sqrt{R^2 - x^2 - y^2}\,.$$



These will serve as the bounds for z! The other two come from the previous analysis of the circle:



Taken together they give us the region:

$$\begin{aligned} U \ = \{(x,y) :\ & -R \le x \le R, \\ & -\sqrt{R^2 - x^2} \le y \le \sqrt{R^2 - x^2}, \\ & -\sqrt{R^2 - x^2 - y^2} \le z \le \sqrt{R^2 - x^2 - y^2}\}\,. \end{aligned}$$

Now the volume is this *triple* integral:

$$\begin{aligned} \text{volume} &= \iiint_W 1\, dV \\ &= \int_{x=-R}^{x=R} \left(\int_{y=-\sqrt{R^2-x^2}}^{y=\sqrt{R^2-x^2}} \left(\int_{z=-\sqrt{R^2-x^2-y^2}}^{z=\sqrt{R^2-x^2-y^2}} 1\, dz \right) dy \right) dx \\ &= \int_{x=-R}^{x=R} \left(\int_{y=-\sqrt{R^2-x^2}}^{y=\sqrt{R^2-x^2}} \left(z \Big|_{z=-\sqrt{R^2-x^2-y^2}}^{z=\sqrt{R^2-x^2-y^2}} \right) dy \right) dx \\ &= \int_{x=-R}^{x=R} \left(\int_{y=-\sqrt{R^2-x^2}}^{y=\sqrt{R^2-x^2}} 2\sqrt{R^2 - x^2 - y^2}\, dy \right) dx\,. \end{aligned}$$

Except for 2, this integral is identical to the integral in the last section.

5.9. *The n-dimensional case*

We outline the properties of integrals in this section.

They are very similar to the ones we for dimensions 1, 2, and 3 that we have seen. In fact, they look simpler because we don't have to repeat the integral signs.

Theorem 5.9.1: Constant Integral Rule For Integrals

Suppose $z = f(X)$ is constant on region U in $\mathbf{R}^n$, i.e., $f(X) = m$ for all X in U and some real number m. Then f is integrable over U and

$$\boxed{\int_U f\,dV = m \cdot V(U)}$$

The interpretation of additivity can be seen as the same as the one for numerical functions: the quantity in a region formed from two regions with a negligible overlap is equal to the sum of the quantities in the two.

- The two intervals overlap by a point, with zero length.
- The two rectangles overlap by an interval, with zero area.
- The two boxes overlap by a rectangle, with zero volume.

There is no double-counting in spite of overlapping!

Theorem 5.9.2: Additivity For Integrals

Suppose $u = f(X)$ is integrable over regions R and over S in $\mathbf{R}^n$ and the two regions overlap only by their boundary points. Then f is integrable $R \cup S$ and we have:

$$\boxed{\int_R f\,dV + \int_S f\,dV = \int_{R\cup S} f\,dV}$$

Warning!

The exact meaning of the boundary is discussed elsewhere.

Theorem 5.9.3: Comparison Rule For Integrals

If

$$f(X) \ge g(X) \text{ on } U\,,$$

for some region U in $\mathbf{R}^n$, then

$$\int_U f\,dV \ge \int_U g\,dV\,,$$

provided $u = f(X)$ and $u = g(X)$ are integrable functions over U. Otherwise we have:

$$\boxed{\begin{array}{lcl} \displaystyle\int_U f\,dV = -\infty & \Longrightarrow & \displaystyle\int_U g\,dV = -\infty \\ \displaystyle\int_U f\,dV = +\infty & \Longleftarrow & \displaystyle\int_U g\,dV = +\infty \end{array}}$$

Theorem 5.9.4: Estimate Rule For Integrals

Suppose $u = f(X)$ is an integrable function over a region U in $\mathbf{R}^n$. Then, if

$$m \le f(X) \le M\,,$$

for all X in U, we have:

$$m \cdot v(U) \le \int_U f\, dV \le M \cdot v(U)$$

Theorem 5.9.5: Constant Multiple Rule For Integrals

Suppose $u = f(X)$ is an integrable function over a region U in $\mathbf{R}^n$. Then so is $c \cdot f$ for any real c and we have:

$$\int_U (c \cdot f)\, dV = c \cdot \int_U f\, dV$$

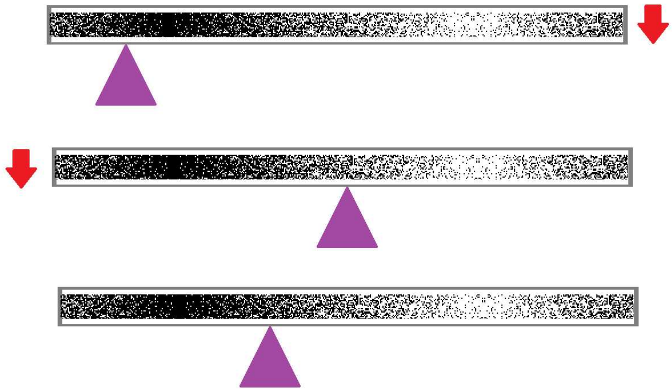
Theorem 5.9.6: Sum Rule For Integrals

Suppose $u = f(X)$ and $u = g(X)$ are integrable functions over region U in $\mathbf{R}^n$. Then so is $f + g$ and we have:

$$\int_U (f+g)\, dV = \int_U f\, dV + \int_U g\, dV$$

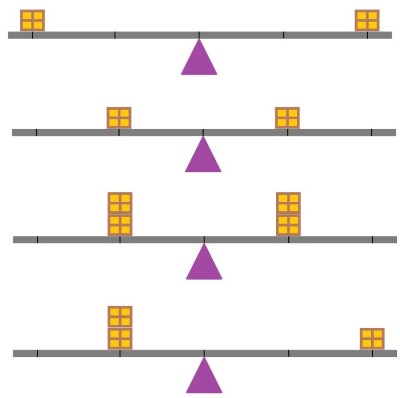
5.10. The center of mass

From Volume 3, we know how to balance this non-uniform rod on a single point of support:



The question is important because this point, called the *center of mass*, is the center of rotation of the object when subjected to a force.

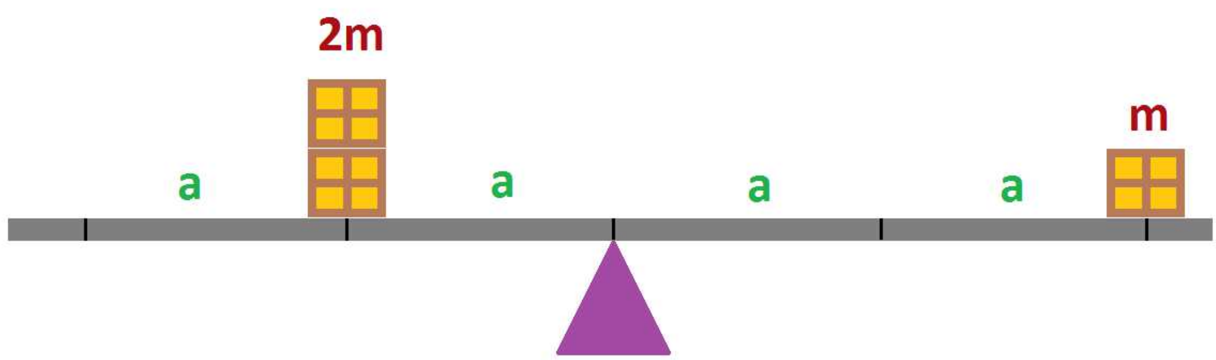
The analysis follows one for the 1-dimensional case; just all the numbers turn into *vectors*!



There are two (or more) axes this time.

This is the 1-dimensional *balance equation*:

$$(a)(2m) = (2a)(m)\,.$$

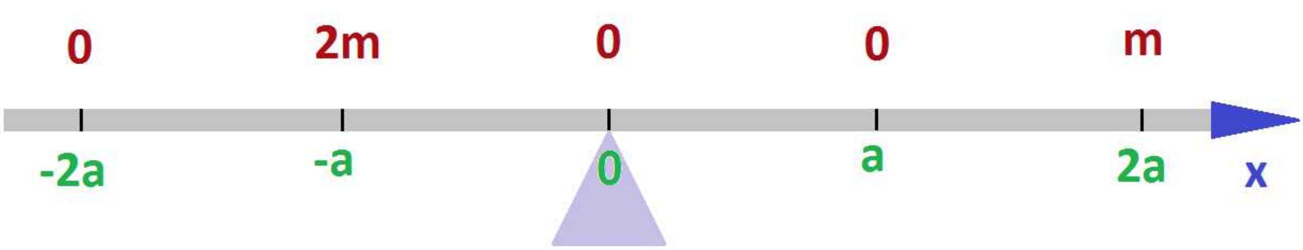


In other words, this expression:

$$\text{distance} \cdot \text{weight}\ ,$$

called *the moment*, is the same to the left and to the right of the support. This distance, also called the *lever*, is now a vector! We simply rewrite the balance equation:

$$(-a)(2m) + (2a)(m) = 0\,.$$



Then,

$$\text{moment} \ = \ \text{vector of location} \cdot \text{weight}\ .$$

Furthermore, we can assume there is an object at every location but the rest of them have 0 mass. The balance equation becomes:

$$... + (-2a)(0) + (-a)(2m) + (0)(0) + (a)(0) + (2a)(m) + ... = 0\,.$$

This analysis bring us to the idea of combining *the weights and the distances* in a proportional manner in order to evaluate the contribution of a particular weight to the overall balance. The balance equation simply says that the sum of all moments is 0:

$$\text{total moment} \ = \sum_i m_i A_i = 0\,,$$

where m_i is the weight of the object located at A_i.

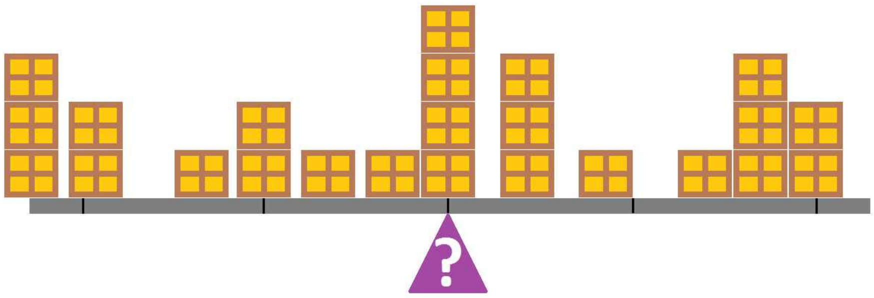
We now go back to the original problem. Suppose different weights are located on a beam, where do we put the support in order to balance it?

It was entirely our decision to place the origin of our coordinate system at the center of mass. The result we have established should be independent from that choice and we can move the origin anywhere. We just need to execute a *change of variables*. Suppose the center of mass (and the origin of the old coordinate system) is located at the point with coordinate Q of the new coordinate system. Then, the new coordinate of the object is

$$C = A + Q\,.$$

Therefore, the balance equation has this form:

$$\sum_C m_C(C - Q) = 0\,.$$



Alternatively, we have:

$$\sum_C m_C C = Q \sum_C m_C\,.$$

It's as if the whole weight is concentrated at Q. Hence the name.

Definition 5.10.1: moment

We can call a *system of weights* a collection of non-negative numbers m_C called *weights* assigned to each point C in a collection of locations in $\mathbf{R}^n$. For a given point Q and for each location C, the scalar product

$$m_C(C - Q)$$

is called the corresponding object's *moment with respect to* Q. The sum of the moments

$$\boxed{\sum_C m_C(C - Q)}$$

is called the *total moment with respect to* Q. The *center of mass of this system of weights* is such a location Q that the total moment with respect to Q is zero; i.e.,

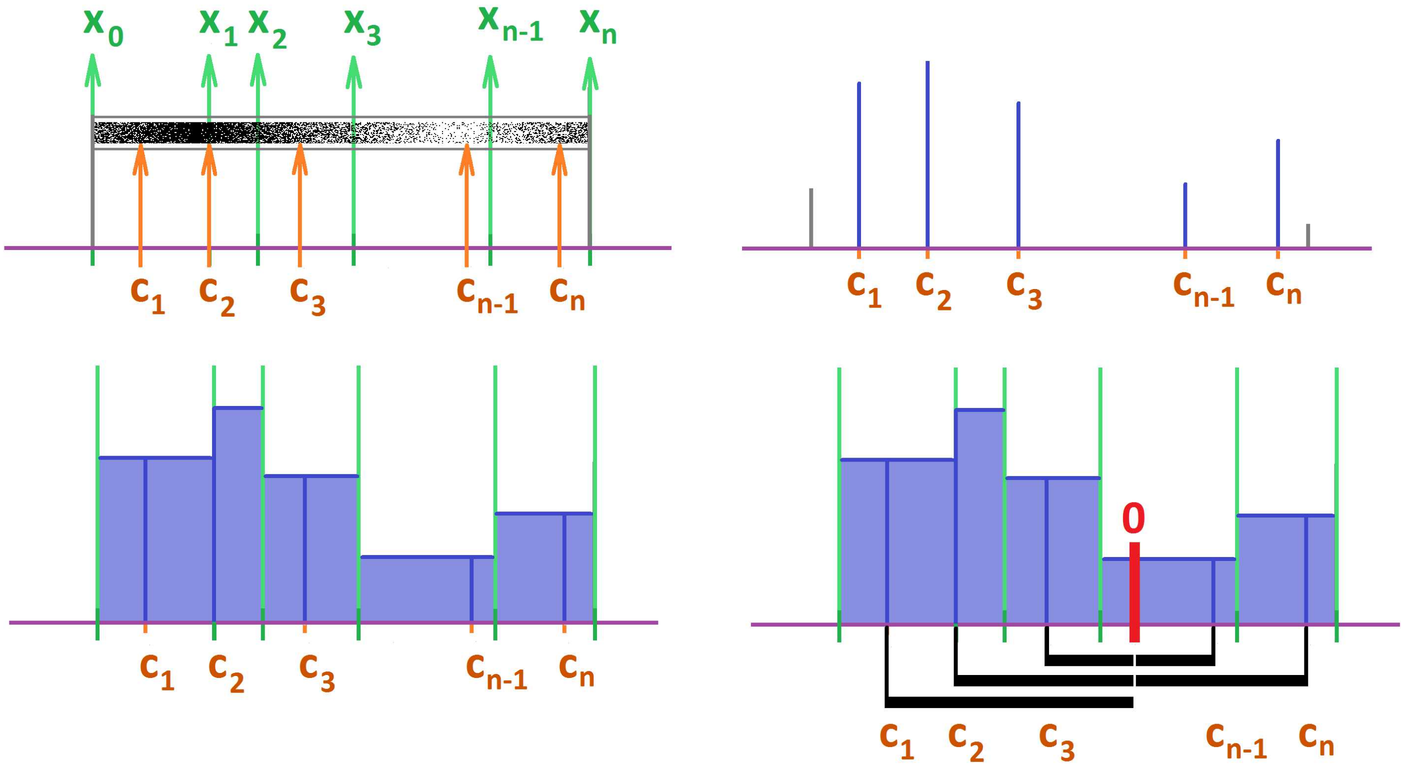
$$\boxed{c = \frac{\sum_C m_C C}{\sum_C m_C}}$$

Exercise 5.10.2

What if we allow the values of m_C to be negative? What is the meaning of the system and of Q?

Suppose an augmented partition P of a n-cell in $\mathbf{R}^n$ is given. Then the density l is known and the terms $l(C)\Delta V_C$ are formed.

The illustration below is for dimension 1:



Each of these terms is a weight shown as a column. The lever of each is also shown as if the weight of the cell is concentrated at point C. Then the total moment of this system of weights with respect to some Q is the Riemann sum,

$$\sum_C m_C(C-Q) = \sum_C l(C)\Delta V(C-Q) = \sum_R G\,\Delta V\,,$$

of the vector-valued function of n variables:

$$G(X) = l(X)(X-Q)\,.$$

Just as above, the system of weights that makes up the object is balanced when the total moment is zero. We arrive to a similar conclusion.

Theorem 5.10.3: Center of Mass Via Sums

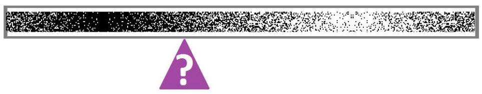
Suppose $z = l(X)$ is a function of n variables defined at points located at the n-cells of a partition of an n-cell R. Then the system of weights $l(X)\Delta X$ has its center of mass at the following point:

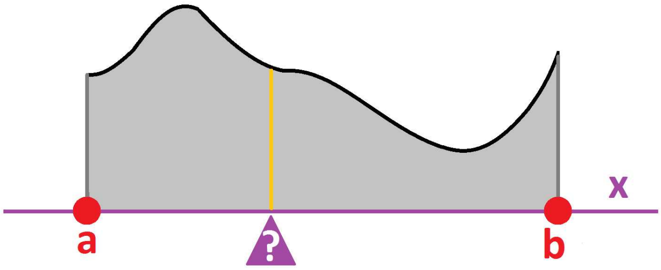
$$\boxed{Q = \frac{\sum_R l(X)X\,\Delta V}{\sum_R l(X)\,\Delta V}}$$

Exercise 5.10.4

Test this formula on some familiar regions.

The next step is to think of the weights assigned to *every* location within R. It's a function!





What we have learned is that the total moment of the region with respect to some Q is approximated by that of this system of weights, which is the Riemann sum,

$$\sum_C m_C(C-Q) = \sum_C l(C_i)\Delta V(C-Q) = \sum_R G\,\Delta V\,,$$

of the vector valued function

$$G(x) = l(X)(X-Q)\,.$$

The system doesn't have to be balanced and the total moment doesn't have to be zero *for each partition*, but it does have to diminish to zero as we refine the partition. This means that the Riemann integral of this function is zero.

Definition 5.10.5: moment

Suppose we have a non-negative function $z=l(X)$ integrable on region R called the *density function*. For a given point C and for each X, the vector-valued function

$$Z = l(X)(X-Q)$$

is called the *moment function with respect to* Q. The integral of the moment function

$$\boxed{\int_R l(X)(X-Q)\,dV}$$

is called the *total moment of the segment with respect to* Q. The *center of mass of the segment* is such a point Q that the total moment with respect to Q is zero.

Theorem 5.10.6: Center of Mass Via Integrals

Suppose we have a non-negative function $z=l(X)$ integrable on region R. If the mass of R is not zero, then the center of mass is located at:

$$\boxed{Q = \frac{\int_R l(X)X\,dV}{\int_R l(X)\,dV}}$$

Proof.

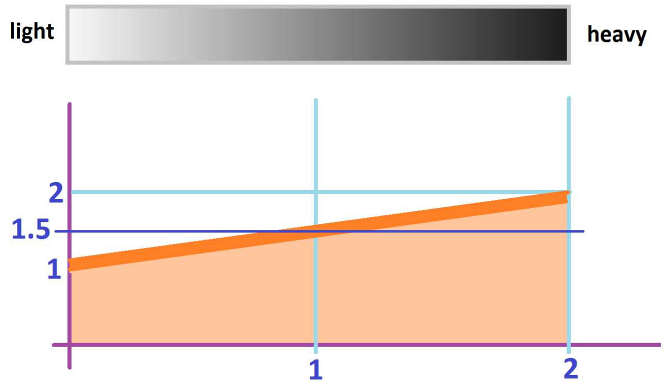
First, we note that $Z=l(X)(X-C)$ is integrable by PR. Then we use SR and CMR to compute the following:

$$0 = \text{ total moment } = \int_R l(X)(X-C)\,dV = \int_R l(X)X\,dV + Q\int_R l(X)\,dV\,.$$

Now solve for C.

Example 5.10.7: linear dependence

Suppose the density of a 2×2 plate R is changing linearly: from 1 to 2. It is similar to this:



Then, $l(x) = x/2 + y/2 + 1$ and the mass is:

$$\begin{aligned} m = \int_R l(X)\, dV &= \int_{[0,1]\times[0,2]} (x/2 + y/2 + 1)\, dV \\ &= \int_0^2 \int_0^2 (x/2 + y/2 + 1)\, dxdy \\ &= \int_0^2 (x^2 + xy/2 + x)\Big|_0^2 dy \\ &= \int_0^2 (2^2 + 2y/2 + 2) - (0^2 + 0y/2 + 0)\, dy \\ &= \int_0^2 (y + 6)\, dy \\ &= (y^2/2 + 6y)\Big|_0^2 dy \\ &= 14\,. \end{aligned}$$

That's the denominator of the fraction. Now, we compute the numerator, the moment:

$$\begin{aligned} M = \int_R l(X)X\, dV &= \int_R (x/2 + y/2 + 1) < x, y > \, dV \\ &= \int_R < (x/2 + y/2 + 1)x \;,\; (x/2 + y/2 + 1)y > \, dV \\ &=< \int_R (x/2 + y/2 + 1)x\, dV \;,\; \int_R (x/2 + y/2 + 1)y\, dV > \\ &= \ldots \end{aligned}$$

Therefore, the center of mass is

$$Q = \frac{M}{m} = \frac{M}{14}\,.$$

Exercise 5.10.8

Finish the problem.

Example 5.10.9: expected value

Let's recall the example of a baker the bread of whose is priced based on the prices of wheat and sugar that change every day. Every day for a month he recorded two numbers – x and y – that represent

how much the two prices deviated from some minimum.

And now he wants to understand what was the *average* combination of prices. For each combination of prices he records how many times is has occurred. He puts these numbers in a table, which makes a *function of two variables*. These may be its inputs and outputs:

$y\backslash x$	0	1	2	...	10
0	$(0,0)$	$(0,1)$	$(0,2)$	...	$(0,10)$
1	$(1,0)$	$(1,1)$	$(2,2)$	...	$(1,10)$
2	$(2,0)$	$(1,1)$	$(2,2)$	...	$(2,10)$
...	...	...	...	...	
10	$(10,0)$	$(10,1)$	$(10,2)$	...	$(10,10)$

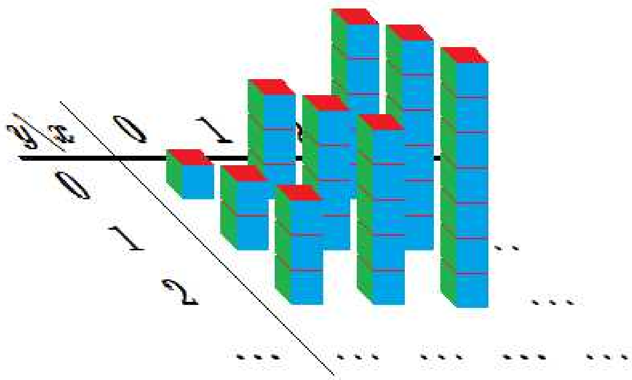
leading to

$y\backslash x$	0	1	2	...	10
0	1	3	5	...	0
1	2	4	6	...	0
2	3	5	7	...	1
...	...	...	...	...	
10	0	1	7	...	0

This may look like a generic function of two variables but let's take a closer look at how the data is collected. It's not the exact value of either price that matters but rather its *range*, say $2 \le x < 3$. This range is an interval of values and together the range of pairs of prices is a *rectangle*, say $[2,3] \times [1,2]$. The data then is represented by a table that looks a bit different:

$y\backslash x$	0		1		2		3	...	9		10
0	•	−−	•	−−	•	−−	•	...	•	−−	•
	|	1	|	3	|	5	|	...	|	0	|
1	•	−−	•	−−	•	−−	•	...	•	−−	•
	|	2	|	4	|	6	|	...	|	0	|
2	•	−−	•	−−	•	−−	•	...	•	−−	•
	|	3	|	5	|	7	|	...	|	1	|
3	•	−−	•	−−	•	−−	•	...	•	−−	•
...		...		...		...		...		...	
9	•	−−	•	−−	•	−−	•	...	•	−−	•
	|	0	|	1	|	7	|	...	|	0	|
10	•	−−	•	−−	•	−−	•	...	•	−−	•

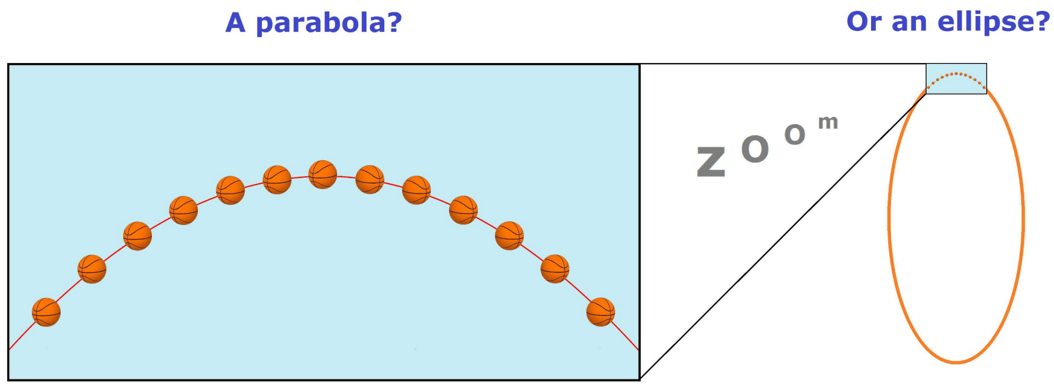
We are justified to visualize this information as columns over these rectangles:



We realize that this isn't just a function of two variables; it's a *discrete* 2-*form*! Furthermore, the average price combination is equivalent – in the dimensions $1,2,3$ – to finding the *center of mass* of this collection of bars.

Example 5.10.10: gravity

A familiar problem about a ball thrown in the air has a solution: its trajectory is a *parabola*. However, we also know that if we throw really-really hard (like a rocket) the ball will start to orbit the Earth following an *ellipse*.



The motion of two planets (or a star and a planet, or a planet and a satellite, etc.) is governed by a single force: the *gravity*. Recall how this force operates.
Newton's Law of Gravity: The force of gravity between two objects is given by the formula:

$$F = G\frac{mM}{r^2}\,,$$

where:

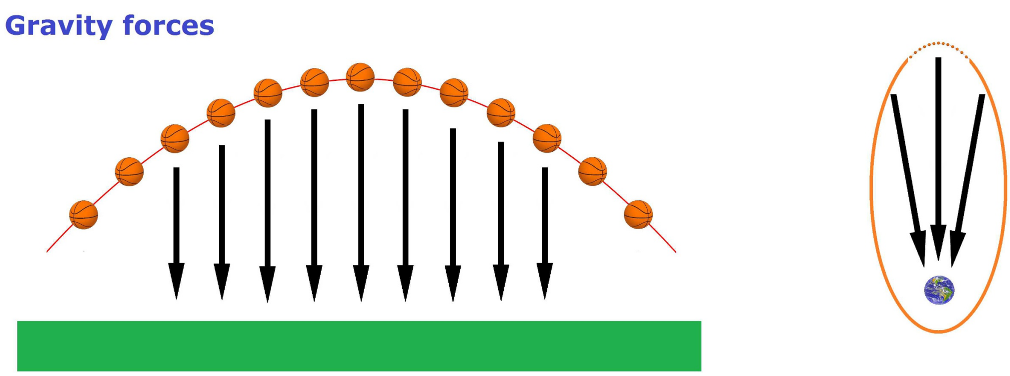
- F is the force between the objects.
- G is the gravitational constant.
- m is the mass of the first object.
- M is the mass of the second object.
- r is the distance between the centers of the masses.

or, in the vector form (with the first object is located at the origin):

$$F = -GmM\frac{X}{||X||^3}\,.$$

This is what we know.

- When the Earth is seen as "large" in comparison to the size of the trajectory, the gravity forces are assumed to be parallel in all locations (the orbit is a parabola).
- When the Earth is seen as "small" in comparison to the size of the trajectory, the gravity forces are assumed to point radially toward that point (the orbit) may be an ellipse, or a hyperbola, or a parabola.



When the size and, therefore, the shape of the Earth matter, things get complicated...

Chapter 6: Vector fields

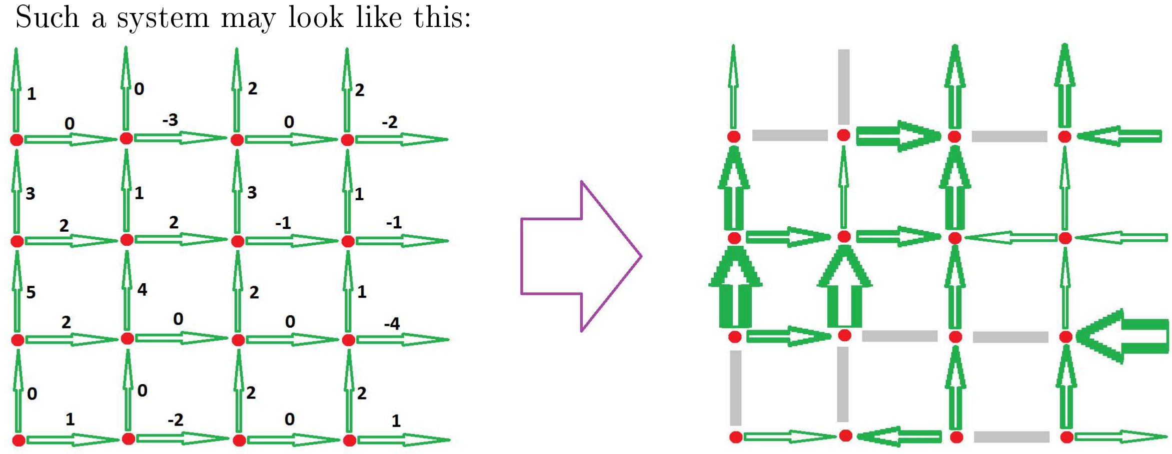
Contents

6.1. What are vector fields?

The first metaphor for a vector field is a *hydraulic system*.

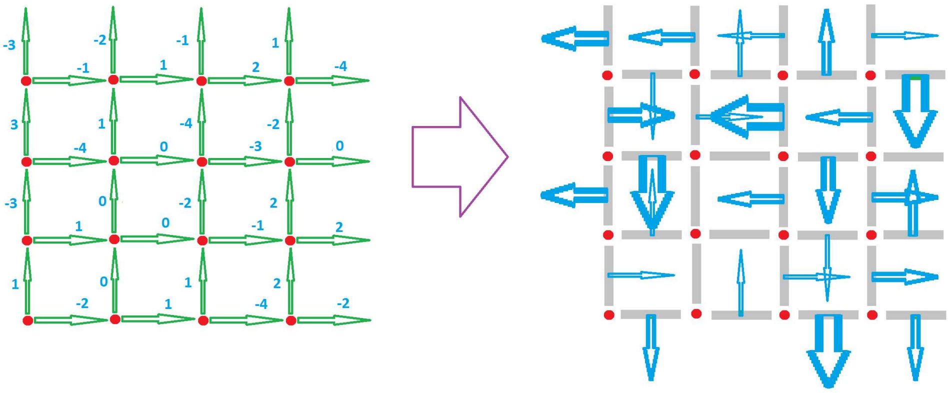
Example 6.1.1: plumbing

Suppose we have a system of pipes with water flowing through them. We model the process with a partition of the plane with its edges representing the pipes and nodes representing the junctions. Then a number is assigned to each edge representing the strength of the flow (in the direction of one of the axes). Such a system may look like this:



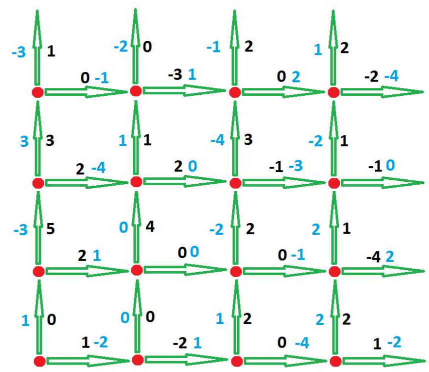
Here the strength of the flow is shown as the thickness of the arrow. This is a real-valued 1-form.

Furthermore, there may be *leakage*. In contrast to the amount of water that actually passes all the way through the pipe, we can make record of the amount that is lost.



That's another real-valued 1-form.

If we assume that the direction of the leakage is perpendicular to the pipe, the two numbers can be combined into a *vector*. The result is a vector-valued 1-form:



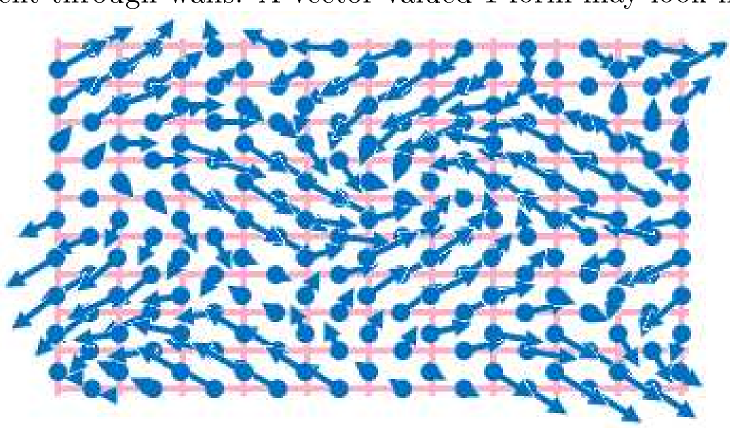
Warning!

The two real-valued 1-forms are re-constructed from the vector-valued 1-form but *not* as its two components but as its projections on the corresponding edges.

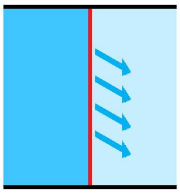
The second metaphor for a vector field is a *flow-through*.

Example 6.1.2: exchange

The data from last example can be used to illustrate a flow of liquid or another material from compartment to compartment through walls. A vector-valued 1-form may look like this:



The situation is reversed in comparison to the last example: the component perpendicular to the edge is the relevant one.

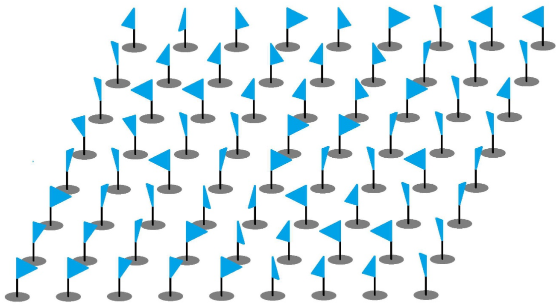


This interpretation is changes in dimension 3 however: the component perpendicular to the *face* is the relevant one.

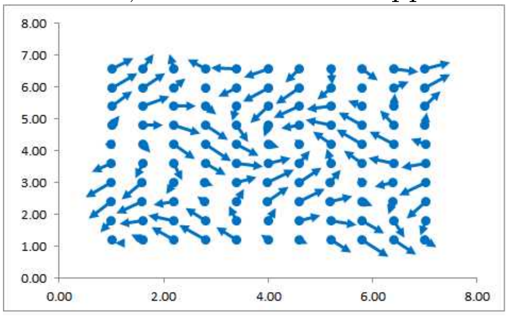
The third metaphor for a vector field is *velocities of particles*.

Example 6.1.3: field of flags

Imagine little flags placed on the lawn; then their directions form a vector field, while the air flow that produced it remains invisible.



Each flag shows the direction (if not the magnitude) of the velocity of the flow at that location. Such flags are also placed on a model airplane in a wind-tunnel. A similar idea is used to model a fluid flow. The dynamics of each particle is governed by the velocity of the flow, at each location, the same at every moment of time. In other words, the vector field supplies a *direction to every location*.



How do we trace the path of a particle? Let's consider this vector field:

$$V(x,y) = < y, -x > \ .$$

Even though the vector field is continuous, the path can be approximated by a parametric curve over a partition of an interval, as follows. At our current location and current time, we examine the vector field to find the velocity and then move accordingly to the next location. We start at this location:

$$X_0 = (2,0)\,.$$

We substitute these two numbers into the equations:

$$V(0,2) = < 2, 0 > \ .$$

This is the direction we will follow. Our next location on the xy-plane is then:

$$X_1 = (0,2) + < 2, 0 > = (2,2)\,.$$

We again substitute these two numbers into V:

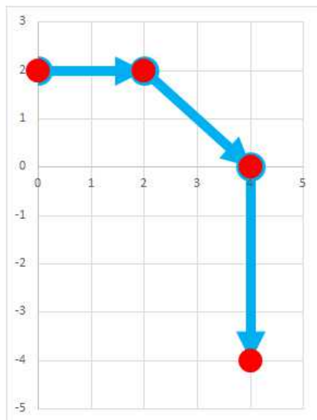
$$V(2,2) = < 2, -2 >\,,$$

leading to the next step. Our next location on the xy-plane is:

$$X_2 = (2, 2)+ < 2, -2 >= (4, 0)\,.$$

One more step: X_2 is substituted into V and our next location is:

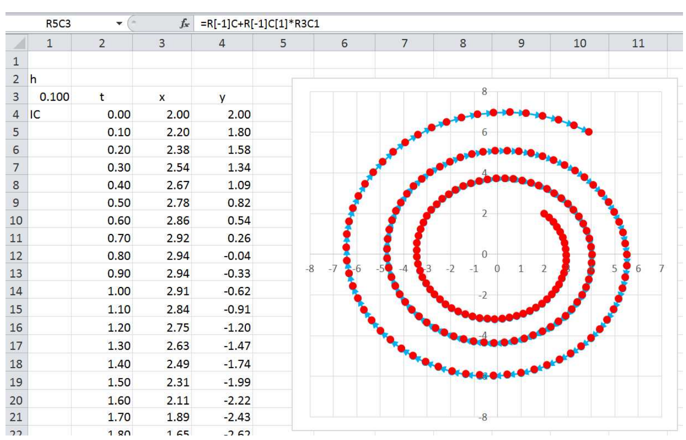
$$X_3 = (4, 0)+ < 0, -4 >= (4, -4)\,.$$



The sequence is spiraling away from the origin. Let's now carry out this procedure with a spreadsheet (with a smaller time increment). The formulas for x_n and y_n are respectively:

`=R[-1]C+R[-1]C[1]*R3C1`   `=R[-1]C-R[-1]C[-1]*R3C1`

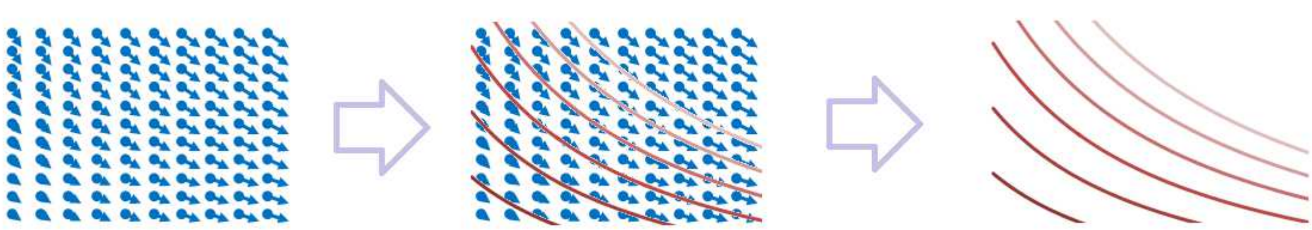
These are the results:



In general, a vector field $V(x, y) =< f(x, y),\ g(x, y) >$ is used to create a system of two *ordinary differential equations* (ODEs):

$$X'(t) = V(X(t)) \quad \text{or} \quad < x'(t),\ y'(t) >= V(x(t),\ y(t)) \quad \text{or} \quad \begin{cases} x'(t) &= f(x(t),\ y(t)),\\ y'(t) &= g(x(t),\ y(t)). \end{cases}$$

Its solution is a pair of functions $x = x(t)$ and $y = y(t)$ that satisfy the equations for every t.



The equations mean that the vectors of the vector field are tangent to these trajectories. ODEs are discussed in Volume 5 (Chapter 5DE-3).

The fourth metaphor for vector fields is a *location-dependent force*.

Example 6.1.4: gravity

Recall that *Newton's Law of Gravity* states that the force of gravity between two objects is given by the formula:

$$f(X) = G\frac{mM}{r^2}\,.$$

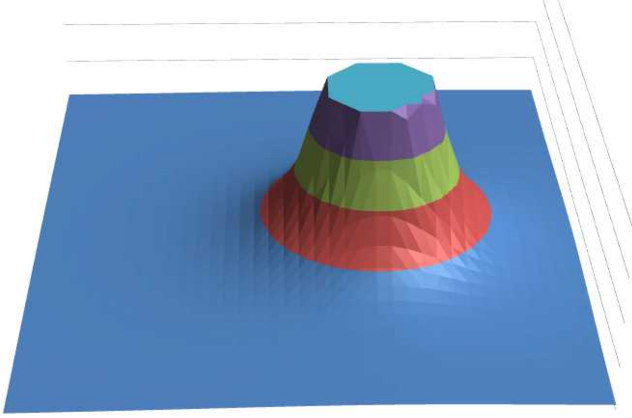
Here:

- f is the *magnitude* of the force between the objects.
- G is the gravitational constant.
- m is the mass of the first object.
- M is the mass of the second object.
- r is the distance between the centers of the masses.

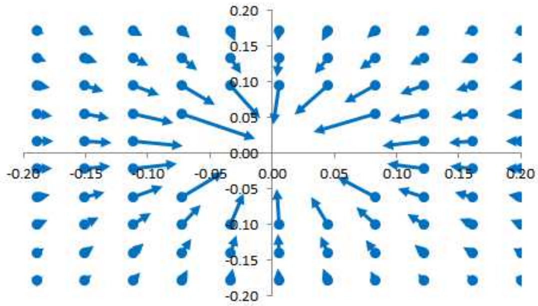
Now, let's assume that the first object is located at the origin. Then the vector of location of the second object is X and the force is a multiple of this vector. If $F(X)$ is the *vector* of the force at the location X, then:

$$F(X) = -GmM\frac{X}{||X||^3}\,.$$

That's the vector form of the law! We plot the magnitude of the force as a function of two variables:



And this is the resulting vector field:



The motion is approximated in the manner described in the last example with the details provided in this chapter.

When the initial velocity of an object is *zero*, it will follow the direction of the force. For example, on object will fall directly on the surface of the Earth. This idea bridges the gap between velocity fields and force fields.

Definition 6.1.5: vector field

A *vector field* is a function defined on a subset of $\mathbf{R}^n$ with values in $\mathbf{R}^n$.

Warning!

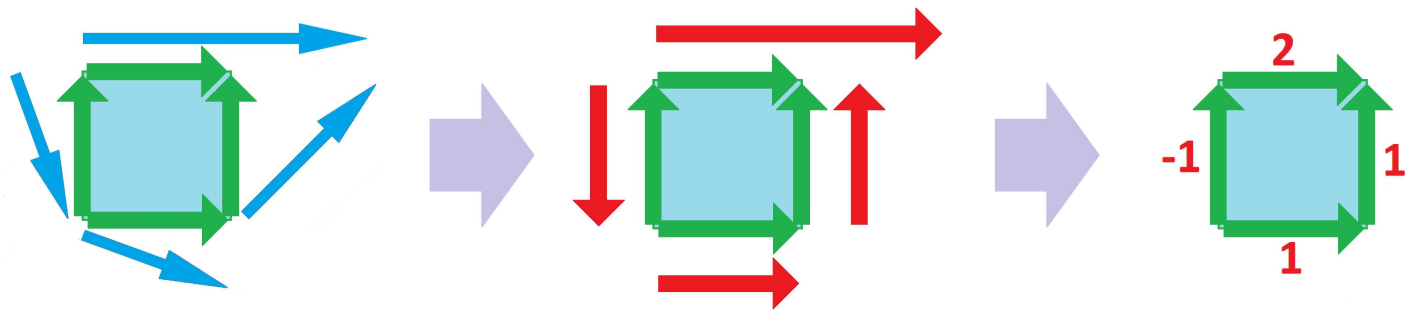
Though unnecessary mathematically, for the purposes of visualization and modeling we think of the input of vector fields as points and outputs as vec-

tors.

But what about the *difference* of a functions of several variables? It's not vector-valued! Some vector fields however might have the difference behind them: the *projection* p of a vector field V on a partition is a function defined at the secondary nodes of the partition as the dot product of the vectors with the corresponding oriented edges:

$$p(C)=V(C)\cdot E\,,$$

where C is the secondary node of the edge E. When the projection of V is the difference of some function, we call V *gradient*.



When no secondary nodes are specified, the formula $p(E)=V(E)\cdot E$ makes a real-valued 1-form from a vector-valued one.

6.2. Motion under forces: a discrete model

Suppose we know the forces affecting a moving object. How can we predict its dynamics?

We simply generalize the 1-dimensional analysis from Volume 2 (Chapter 2DC-6) to the vector case.

Assuming a fixed mass, the total force gives us our acceleration. We are to compute:

- the velocity from the acceleration, and then
- the location from the velocity.

A fixed time increment Δt is supplied ahead of time even though it can also be variable.

We start with the following three quantities that come from the setup of the motion:

- the initial time t_0,
- the initial velocity V_0, and
- the initial location P_0.

They are placed in the consecutive cells of the first row of the spreadsheet:

	iteration n	time t_n	acceleration A_n	velocity V_n	location P_n
initial:	0	3.5	$--$	$<33,44>$	$<22,11>$

As we progress in time and space, new numbers are placed in the next row of our spreadsheet. There is a *set of columns* for each vector, two or three depending on the dimension.

Just as before, we rely on *recursive formulas*.

The current acceleration A_0 given in the first cells of the second row. The current velocity V_1 is found and placed in the second pair (or triple) of cells of the second row of our spreadsheet:

- current velocity = initial velocity + current acceleration · time increment. The second quantity we use is the initial location P_0. The following is placed in the third set of cells of the second row:
- current location = initial location + current velocity · time increment.

This dependence is shown below:

	iteration n	time t_n	acceleration A_n		velocity V_n		location P_n
initial:	0	3.6	$--$		$< 33, 44 >$		$< 22, 11 >$
					$\downarrow$		$\downarrow$
current:	1	t_1	$< 66, 77 >$	$\rightarrow$	V_1	$\rightarrow$	P_1

We continue with the rest in the same manner. As we progress in time and space, numbers and vectors are supplied and placed in each of the four sets of columns of our spreadsheet one row at a time:

$$t_n,\ A_n,\ V_n,\ P_n,\ n = 1, 2, 3, \ldots$$

The first quantity in each row we compute is the time:

$$t_{n+1} = t_n + \Delta t\,.$$

The next is the acceleration A_{n+1}. Where does it come from? It may come as pure data: the column is filled with number ahead of time or it is being filled as we progress in time and space. Alternatively, there is an explicit, functional dependence of the acceleration (or the force) on the rest of the quantities. The acceleration may be a function of the following:

- the current time, e.g., $A_{n+1} =< \sin t_{n+1},\ \cos t_{n+1} >$, such as when we speed up the car, or
- The last location, such as when the gravity depends on the distance to the planet (below), or
- The last velocity, e.g., $A_{n+1} = -V_n$ such as when the air resistance works in the opposite direction of the velocity,

or all three.

The nth iteration of the velocity V_n is computed:

- current velocity = last velocity + current acceleration · time increment,
- $V_{n+1} = V_n + A_n \cdot \Delta t$.

The values of the velocity are placed in the second set of columns of our spreadsheet.

The nth iteration of the location P_n is computed:

- current location = last location + current velocity · time increment,
- $P_{n+1} = P_n + V_n \cdot \Delta t$.

The values of the location are placed in the third set of columns of our spreadsheet.

The result is a growing table of values:

	iteration n	time t_n	acceleration A_n	velocity V_n	location P_n
initial:	0	3.5	$--$	$< 33, 44 >$	$< 22, 11 >$
	1	3.6	$< 66, 77 >$	$< 38.5, 45.1 >$	$< 25.3, 13.0 >$
	...	...	...	...	...
	1000	103.5	$< 666, 777 >$	$< 4, 1 >$	$< 336, 200 >$
	...	...	...	...	...

The result may be seen as four sequences t_n, A_n, V_n, P_n or as the table of values of three *vector-valued functions* of t.

Exercise 6.2.1

Implement a variable time increment: $\Delta t_{n+1} = t_{n+1} - t_n$.

Example 6.2.2: rolling ball

A rolling ball is unaffected by horizontal forces. Therefore, $A_n = 0$ for all n. The recursive formulas for the horizontal motion simplify as follows:

- The velocity $V_{n+1} = V_n + A_n \cdot \Delta t = V_n = V_0$ is constant.
- The position $P_{n+1} = P_n + V_n \cdot \Delta t = P_n + V_0 \cdot \Delta t$ grows at equal increments.

In other words, the position depends linearly on the time.

Example 6.2.3: falling ball

A falling ball is unaffected by horizontal forces and the vertical force is constant: $A_n = A$ for all n. The first of the two recursive formulas for the vertical motion simplifies as follows:

- The velocity $V_{n+1} = V_n + A_n \cdot \Delta t = V_n + A \cdot \Delta t$ grows at equal increments.
- The position $P_{n+1} = P_n + V_n \cdot \Delta t$ grows at linearly increasing increments.

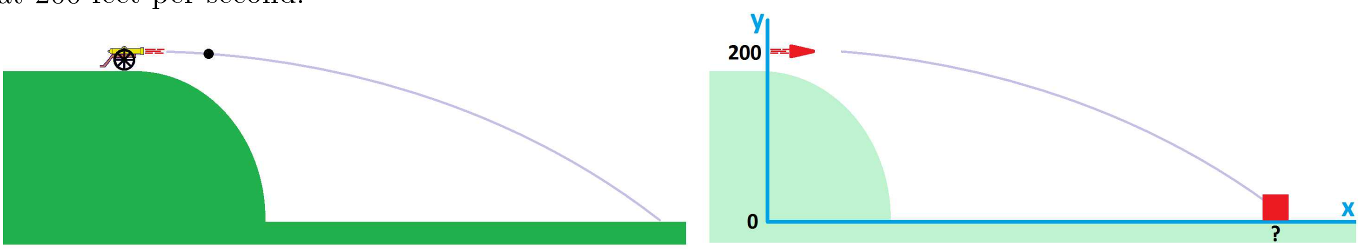
In other words, the position depends quadratically on the time.

Example 6.2.4: vector algebra

A falling ball is unaffected by horizontal forces and the vertical force is constant:

$$A_n =< 0, -g > \ .$$

Now recall the setup considered previously: from a 200 feet elevation, a cannon is fired horizontally at 200 feet per second.



The initial conditions are:

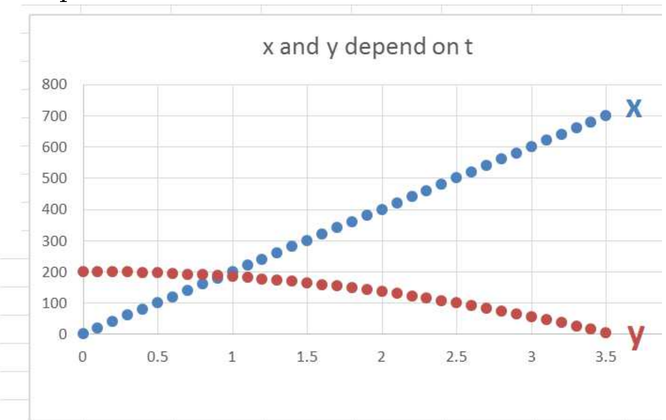
- the initial location, $P_0 =< 0, 200 >$,
- the initial velocity, $V_0 =< 200, 0 >$.

Then we have recursive vector equations:

$$V_{n+1} = V_n + < 0, -g > \Delta t \quad \text{and} \quad P_{n+1} = P_n + V_n \Delta t \,.$$

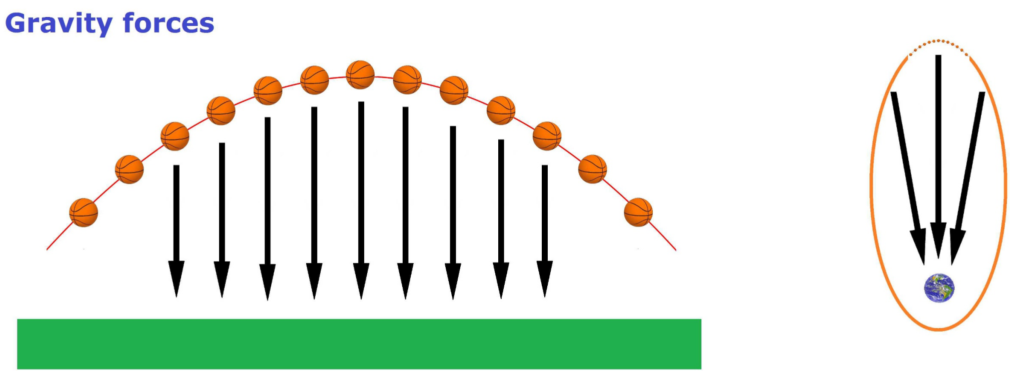
Implemented with a spreadsheet, the formulas produce these results:

time	depth x	height y
0	0	200
0.1	20	199.84
0.2	40	199.36
0.3	60	198.56
0.4	80	197.44
0.5	100	196
0.6	120	194.24
....		
3.2	640	36.16
3.3	660	25.76
3.4	680	15.04
3.5	700	$\boxed{4}$
3.6	720	$\boxed{-7.36}$

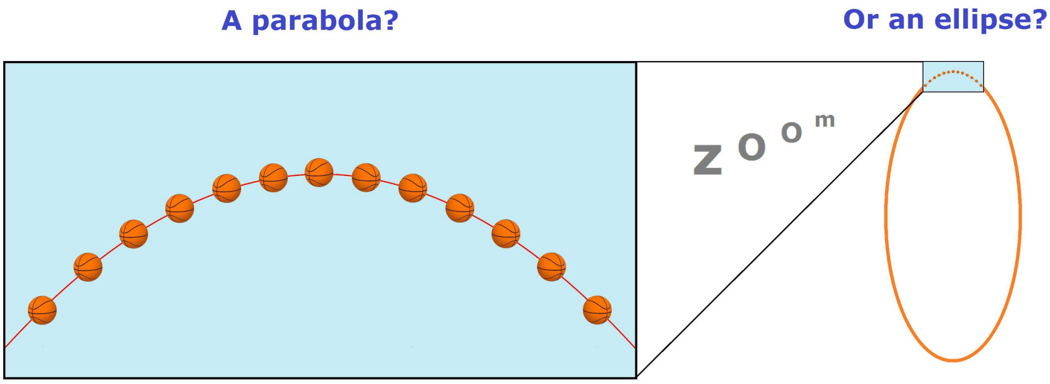


Example 6.2.5: planets

Let's apply what we have learned to *planetary motion*. The problem above about a ball thrown in the air has a solution: its trajectory is a *parabola*.



However, we also know that if we throw really-really hard (like a rocket), the ball will start to orbit the Earth following an *ellipse*.

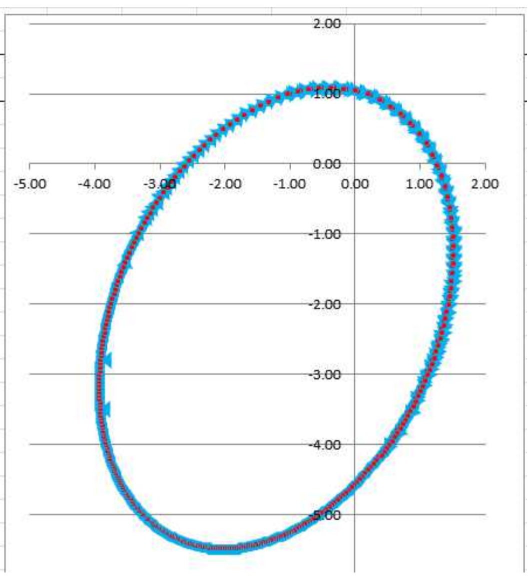


The motion of two planets (or the sum and a planet, or a planet and a satellite, etc.) is governed by the *Newton Law of Gravity*. From this law, another law of motion can be derived. Consider the *Kepler's Laws of Planetary Motion:*

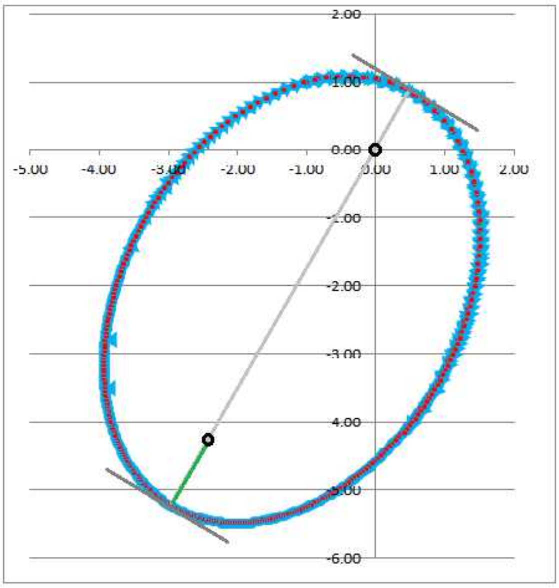
- The orbit of a planet is an ellipse with the Sun at one of the two foci.
- A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.
- The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

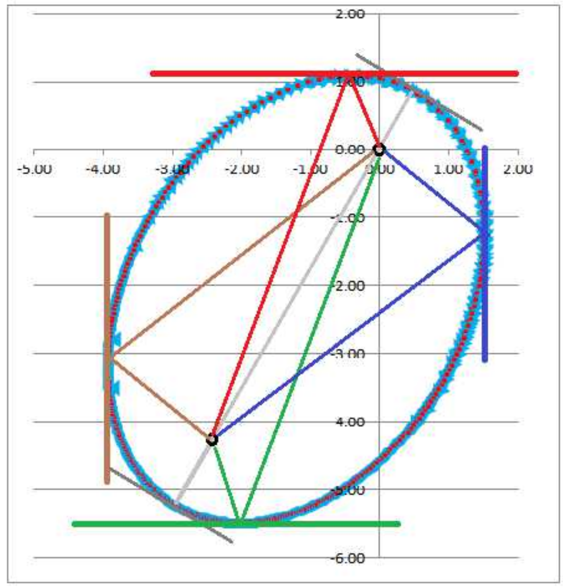
To confirm the law, we use the formulas above but this time the acceleration depends on the location, as follows:

		X''=-X/|X|^3											
0.15										Conservation			Kepler II
	time	acceleration		velocity		speed	position		distance	kinetic	potential	total	area
n	t	x''	y''	x'	y'	s	x	y	D	energy	energy		swept
0	0.00			1.00	0.30	1.04	-1.00	1.00	1.41	0.545	0.000	0.545	
1	0.15	0.4	-0.4	1.05	0.25	1.08	-0.84	1.04	1.34	0.585	-0.040	0.545	0.098
2	0.30	0.4	-0.4	1.11	0.18	1.12	-0.68	1.06	1.26	0.628	-0.083	0.545	0.098
3	0.45	0.3	-0.5	1.16	0.10	1.16	-0.50	1.08	1.19	0.674	-0.129	0.545	0.098
4	0.60	0.3	-0.6	1.20	0.01	1.20	-0.32	1.08	1.13	0.722	-0.177	0.545	0.098
5	0.75	0.2	-0.8	1.23	-0.11	1.24	-0.14	1.06	1.07	0.768	-0.223	0.545	0.098
6	0.90	0.1	-0.9	1.25	-0.24	1.27	0.05	1.03	1.03	0.811	-0.266	0.545	0.098
7	1.05	0.0	-0.9	1.24	-0.38	1.30	0.24	0.97	1.00	0.846	-0.301	0.545	0.098
8	1.20	-0.2	-1.0	1.21	-0.52	1.32	0.42	0.89	0.99	0.868	-0.323	0.545	0.098
9	1.35	-0.4	-0.9	1.14	-0.66	1.32	0.59	0.79	0.99	0.873	-0.328	0.545	0.098
10	1.50	-0.6	-0.8	1.05	-0.78	1.31	0.75	0.68	1.01	0.862	-0.317	0.545	0.098
11	1.65	-0.7	-0.7	0.94	-0.88	1.29	0.89	0.54	1.04	0.836	-0.291	0.545	0.098
12	1.80	-0.8	-0.5	0.83	-0.96	1.26	1.01	0.40	1.09	0.798	-0.253	0.545	0.098
13	1.95	-0.8	-0.3	0.71	-1.00	1.23	1.12	0.25	1.15	0.753	-0.208	0.545	0.098
14	2.10	-0.7	-0.2	0.60	-1.03	1.19	1.21	0.10	1.21	0.706	-0.161	0.545	0.098
15	2.25	-0.7	-0.1	0.50	-1.04	1.15	1.28	-0.06	1.29	0.659	-0.114	0.545	0.098
16	2.40	-0.6	0.0	0.41	-1.03	1.11	1.34	-0.21	1.36	0.614	-0.069	0.545	0.098
17	2.55	-0.5	0.1	0.33	-1.02	1.07	1.39	-0.37	1.44	0.571	-0.026	0.545	0.098
18	2.70	-0.5	0.1	0.26	-1.00	1.03	1.43	-0.52	1.52	0.533	0.012	0.545	0.098
19	2.85	-0.4	0.1	0.19	-0.98	1.00	1.46	-0.66	1.60	0.497	0.048	0.545	0.098

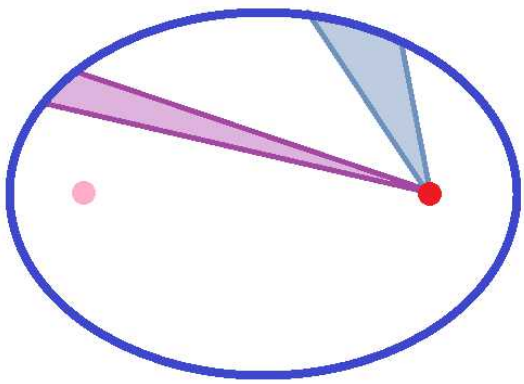


The resulting trajectory does seem to be an ellipse (confirmed by finding its foci):



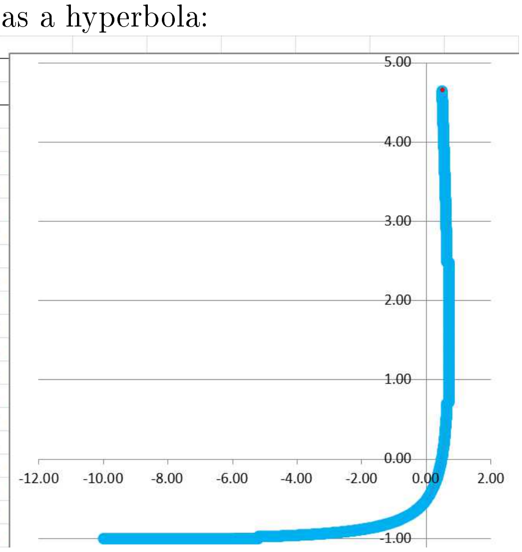


Note that the Second Kepler's Law implies that the motion is different from one provided by the standard parametrization of the ellipse.



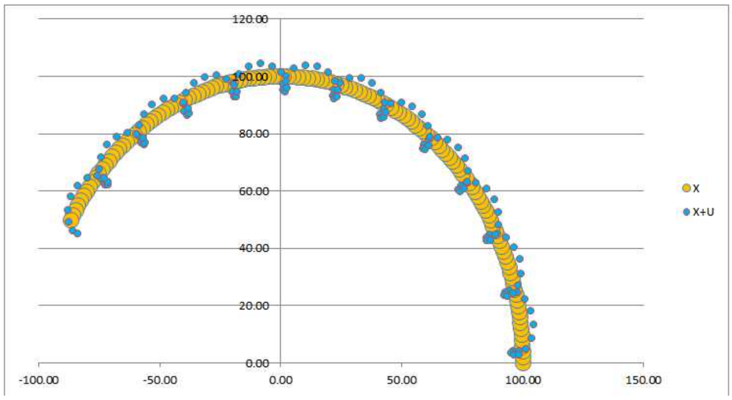
Our computation can produce other kinds of trajectories such as a hyperbola:

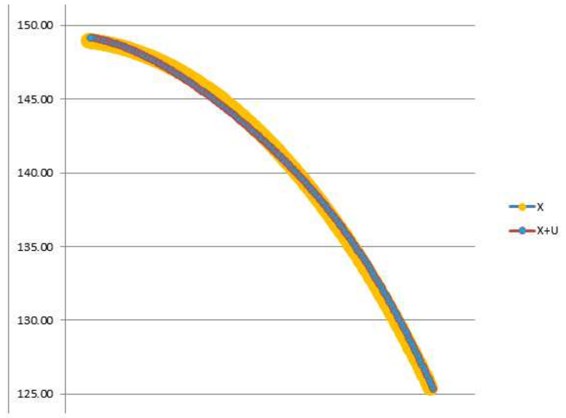
0.01	3.0	X''=-X/|X|^m								Conservation			Kepler II
	time	acceleration		velocity		speed	position		distance	kinetic	potential	total	area
n	t	x''	y''	x'	y'	s	x	y		energy	energy		swept
0	0.00			1.00	0.00	1.00	-10.00	-1.00	10.05	0.500	0.000	0.500	
1	0.01	0.0	0.0	1.00	0.00	1.00	-9.99	-1.00	10.04	0.500	0.000	0.500	0.005000
2	0.02	0.0	0.0	1.00	0.00	1.00	-9.98	-1.00	10.03	0.500	0.000	0.500	0.005000
3	0.03	0.0	0.0	1.00	0.00	1.00	-9.97	-1.00	10.02	0.500	0.000	0.500	0.005000
4	0.04	0.0	0.0	1.00	0.00	1.00	-9.96	-1.00	10.01	0.500	0.000	0.500	0.005000
5	0.05	0.0	0.0	1.00	0.00	1.00	-9.95	-1.00	10.00	0.500	0.000	0.500	0.005000
6	0.06	0.0	0.0	1.00	0.00	1.00	-9.94	-1.00	9.99	0.501	-0.001	0.500	0.005000
7	0.07	0.0	0.0	1.00	0.00	1.00	-9.93	-1.00	9.98	0.501	-0.001	0.500	0.005000
8	0.08	0.0	0.0	1.00	0.00	1.00	-9.92	-1.00	9.97	0.501	-0.001	0.500	0.005000
9	0.09	0.0	0.0	1.00	0.00	1.00	-9.91	-1.00	9.96	0.501	-0.001	0.500	0.005000
10	0.10	0.0	0.0	1.00	0.00	1.00	-9.90	-1.00	9.95	0.501	-0.001	0.500	0.005000
11	0.11	0.0	0.0	1.00	0.00	1.00	-9.89	-1.00	9.94	0.501	-0.001	0.500	0.005000
12	0.12	0.0	0.0	1.00	0.00	1.00	-9.88	-1.00	9.93	0.501	-0.001	0.500	0.005000
13	0.13	0.0	0.0	1.00	0.00	1.00	-9.87	-1.00	9.92	0.501	-0.001	0.500	0.005000
14	0.14	0.0	0.0	1.00	0.00	1.00	-9.86	-1.00	9.91	0.501	-0.001	0.500	0.005000
15	0.15	0.0	0.0	1.00	0.00	1.00	-9.85	-1.00	9.90	0.501	-0.001	0.500	0.005000
16	0.16	0.0	0.0	1.00	0.00	1.00	-9.84	-1.00	9.89	0.502	-0.002	0.500	0.005000
17	0.17	0.0	0.0	1.00	0.00	1.00	-9.83	-1.00	9.88	0.502	-0.002	0.500	0.005000
18	0.18	0.0	0.0	1.00	0.00	1.00	-9.82	-1.00	9.87	0.502	-0.002	0.500	0.005000



Example 6.2.6: Sun-Earth-Moon

The Earth revolves around the Sun and the Moon revolves around the Earth. The result derived from such a generic description should look like the one on left.





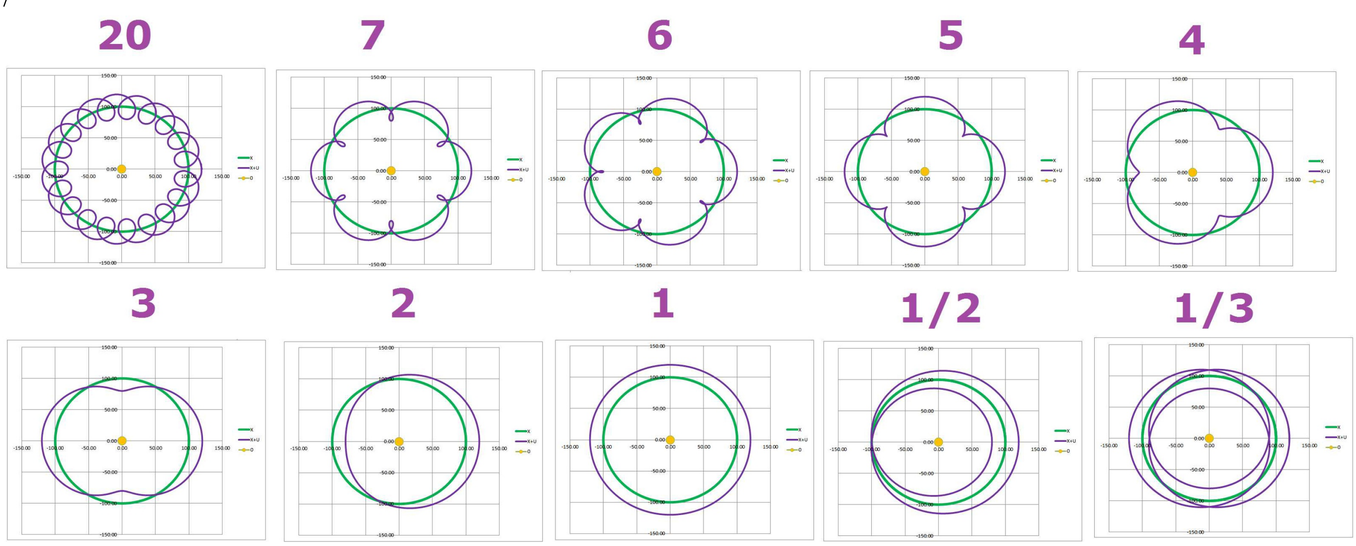
Now, let's use the actual data:

- The average distance between the Earth and the Sun is 149.60 million km.
- The average distance between the Moon and the Earth is $385,000$ km.
- The Moon orbits Earth one revolution in 27.323 days.

The paths are plotted on right. As you can see, not only the Moon never goes backwards but also its orbit is in fact convex! (By "convex orbit" we mean "convex region inside the orbit": any two points inside are connected by the segment that is also inside.)

Example 6.2.7: simulations

Below we have: a hypothetical star (orange) is orbited by a planet (blue) which is also orbited by its moon (purple). Now we vary the number of times per year the moon orbits the planet, from 20 to 1/3.

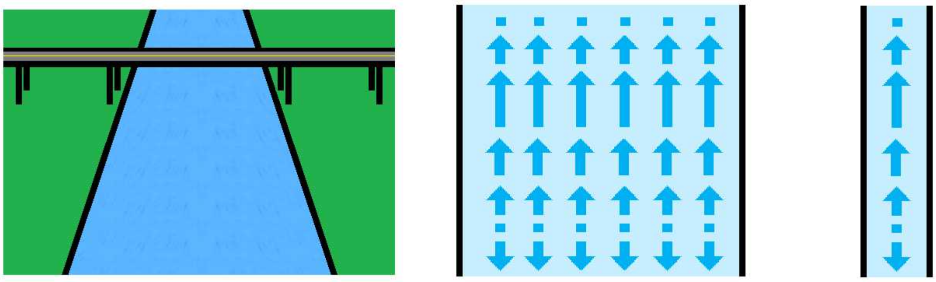


6.3. *The algebra and geometry of vector fields*

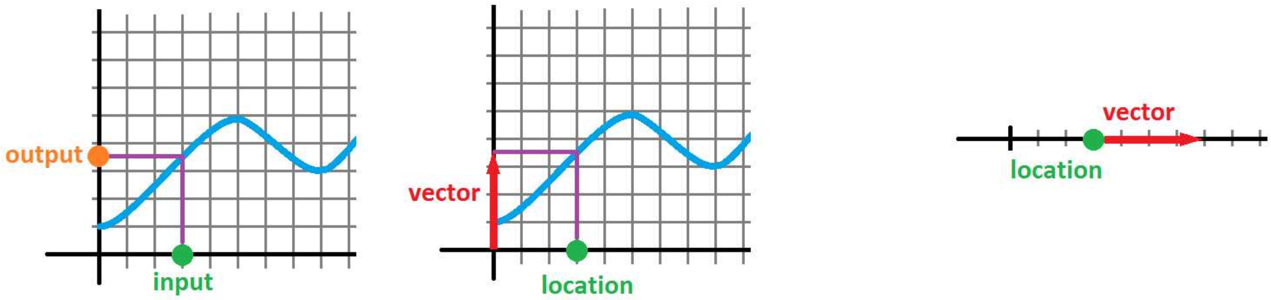
Vector fields appear in all dimensions. The idea is the same: there is a flow of liquid or gas and we record how fast a single particle at every location is moving.

Example 6.3.1: dimension 1

The flow is in a pipe. The same idea applies to a canal with the water that has the exact same velocity at all locations across it.



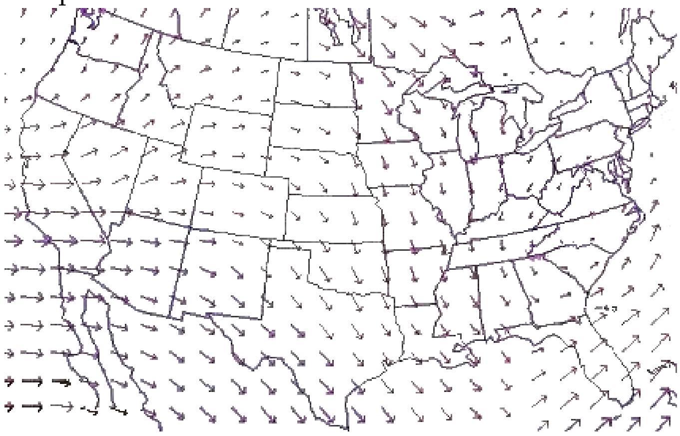
Of course these are just numerical functions:



This is just another way to visualize them.

Example 6.3.2: dimension 2

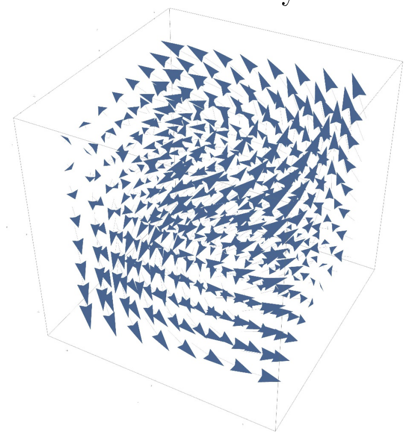
Not every vector field of dimension $n > 1$ is gradient and, therefore, some of them cannot be visualized as flows on a surface under nothing but gravity. A vector field of dimension $n = 2$ is then seen as a flow on the plane: liquid in a pond or the air over a surface of the Earth.



The metaphor applies under the assumption that the air or water has the exact same velocity at every locations regardless of the elevation.

Example 6.3.3: dimension 3

This time, a vector field is thought of as a flow without any restrictions on the velocities of the particles.

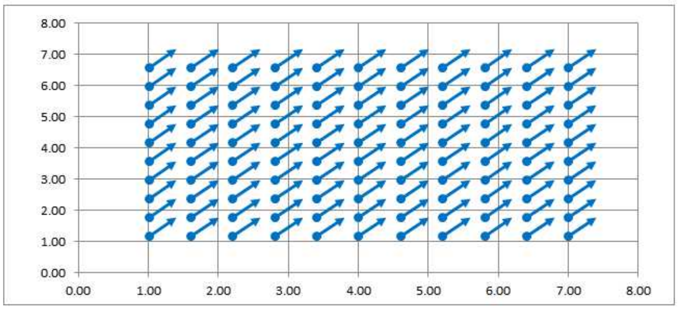


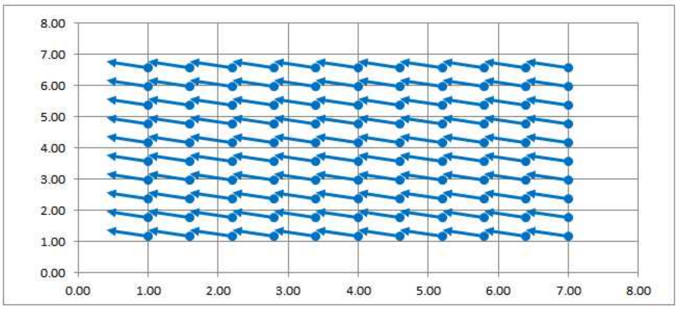
A model of stock prices as a flow will lead to 10,000-dimensional vector field. This necessitates our use of *vector notation*. We also start thinking of the input, just as the output, to be vectors (of the same dimension). For example, the two "radial" vector fields in the last section have the same representation:

$$V(X) = X\,.$$

An even simpler vector field is a constant:

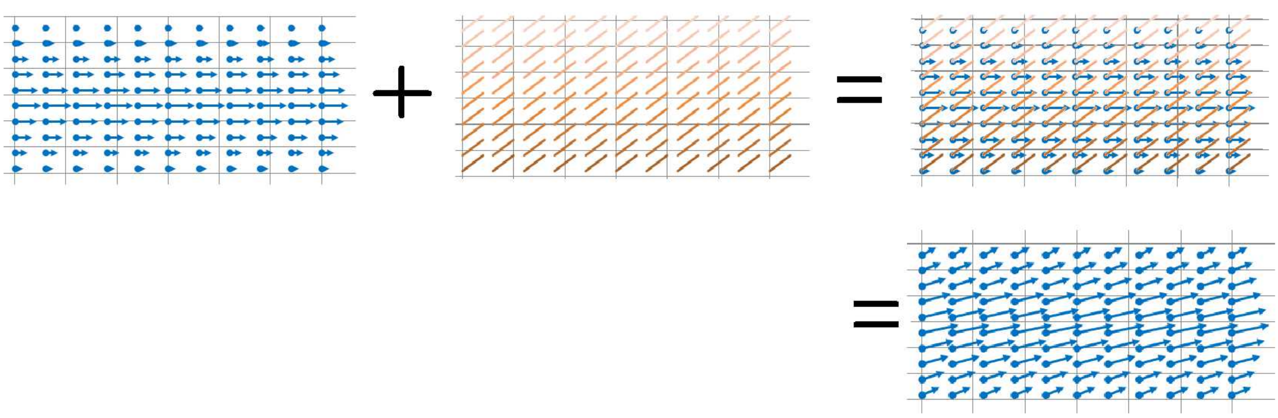
$$V(X) = V_0\,.$$





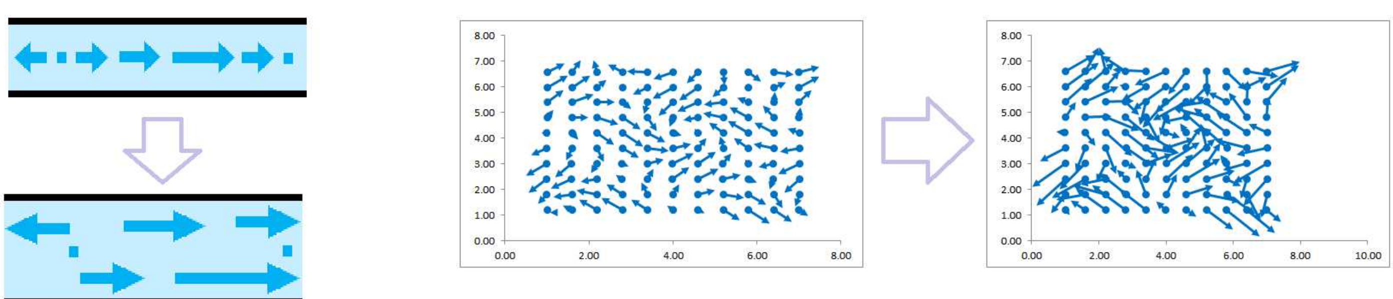
Each vector field is just a vector – at a fixed location. Then it is just a location-dependent (but time-independent!) vector but still a vector. That is why all algebraic operation for vectors are applicable to vector fields.

First, *addition*. Imagine that that we have a river – with the velocities of the water particles represented by vector field V – and then wind starts – with the velocities of the air particles represented by vector field U. This is their sum:

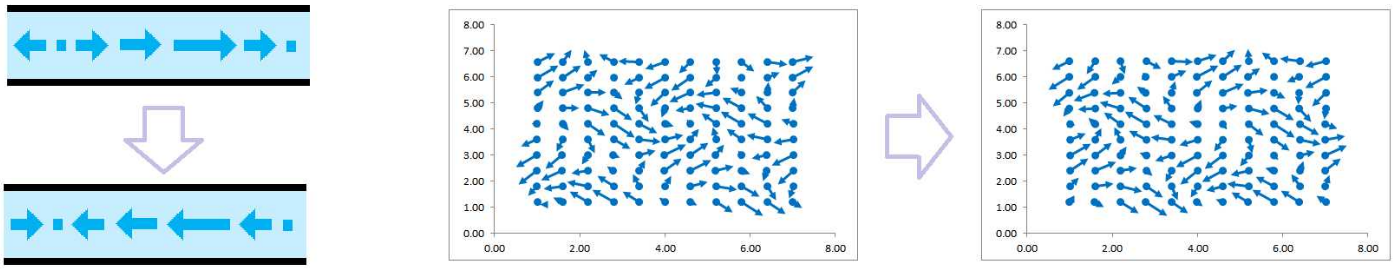


One can argue that the resulting dynamics of water particles will be represented by the vector field $V + U$.

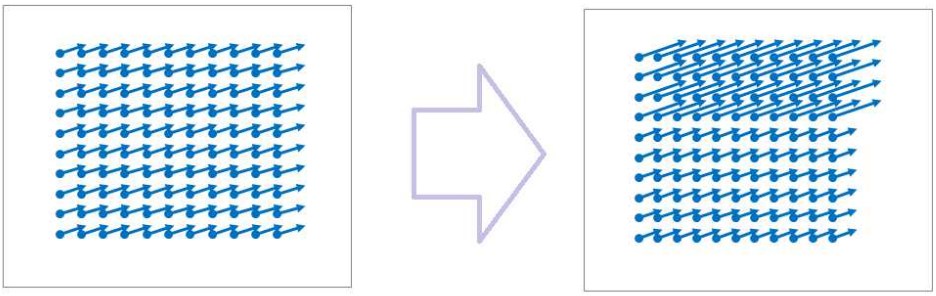
Second, *scalar multiplication*. If the velocities of the water particles in a pipe are represented by vector field V and we then double the pressure (i.e., pump twice as much water), we expect the new velocities to be represented by the vector field $2V$.



Reversing the flow will be represented by the vector field $-V$.



Furthermore, the scalar might also be location-dependent, i.e., we are multiplying our vector field in $\mathbf{R}^n$ by a (scalar) function of n variables.



The computations with specific vector fields are carried out

- one location at a time and
- one component at a time.

Now *geometry*.

What is the magnitude of a vector? As a function, it takes a vector as an input and produces a number as the output. It's just another function of n variables. We can apply it to vector fields, producing this composition:

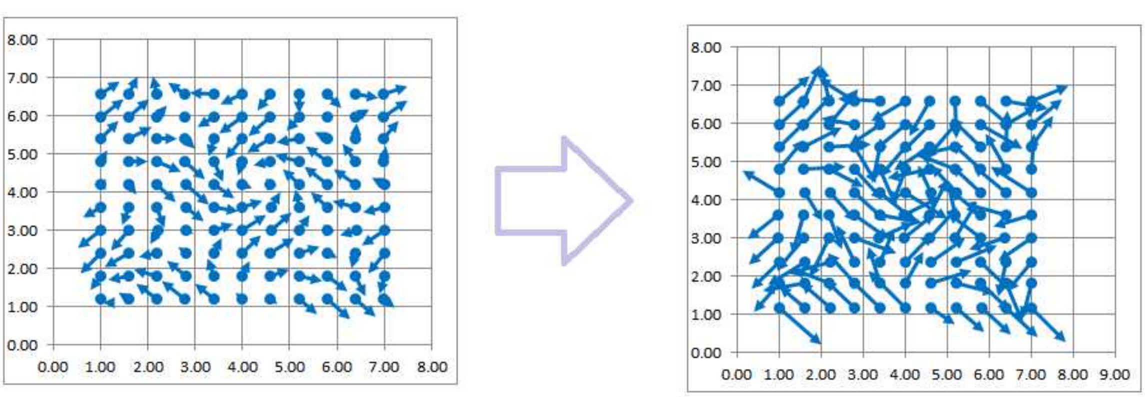
$$\boxed{f(X) = ||V(X)||}$$

The result is a function of n variables that gives us the magnitude of the vector $V(X)$ at location X. The construction is exemplified by the "scalar" version of the Newton's Law of Gravity.

Furthermore, we can use this function to modify vector fields in a special way:

$$\boxed{W(X) = \frac{V(X)}{||V(X)||}}$$

The result is a new vector fields with the exactly same directions of the vectors but with *unit length* (unless it's zero).



This construction is called *normalization*.

Warning!

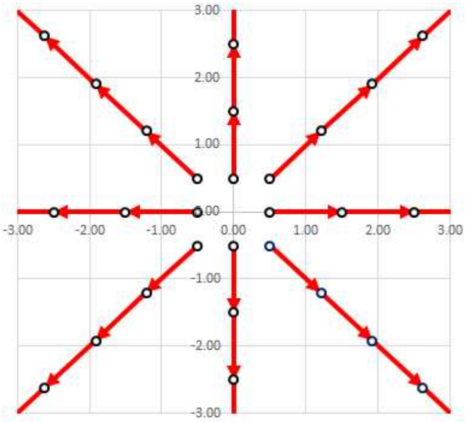
The domain of the new vector field might change as it is undefined at those X where $V(X) = 0$.

Example 6.3.4: normalized outflow

The "accelerated outflow" presented in the first section is no longer accelerated after normalization:

$$W(X) = \frac{X}{||X||}\,.$$

The speed is constant!



The price we pay for making the vector field well-behaved is the appearance of a hole in the domain, $X \neq 0$.

Exercise 6.3.5

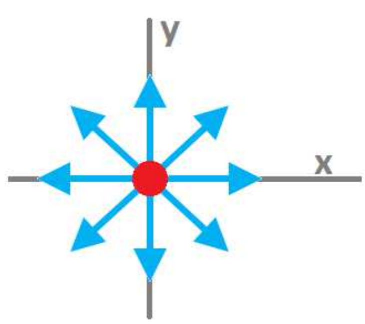
Show that the hole can't be repaired, in the following sense: there is no such vector U that $||W(X) - U|| \to 0$ as $X \to 0$ (i.e., this is a non-removable discontinuity).

Exercise 6.3.6

What if we do the dot product of two vector fields?

If can we rotate a vector, we can rotate vector fields V? In dimension 2, the normal vector field of a vector field $V =< u, v >$ on the plane is given by

$$V^\perp =< u, v >^\perp =< -v, u >\ .$$

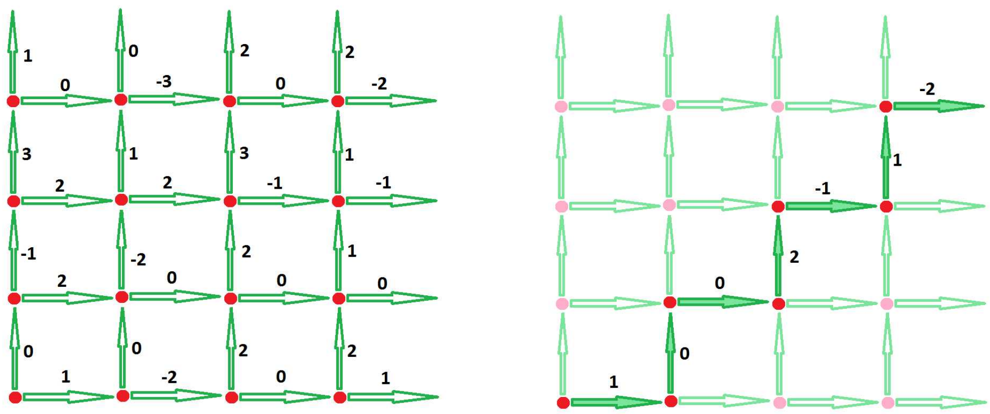


We have then a special operation on vectors fields. For example, rotating a constant vector field is also constant. However, the normal of the rotation vector field is the radial vector field.

6.4. Summation along a curve: flow and work

Example 6.4.1: pipes

We look at this as a system of *pipes* with the numbers indicating the rate of the flow in each pipe (along the directions of the axes):



What is the total flow along this "staircase"? We simply add the values located on these edges:

$$W = 1 + 0 + 0 + 2 + (-1) + 1 + (-2)\,.$$

But these edges just happen to be positively oriented. What if we, instead, go around the first square? We have the following:

$$W = 1 + 0 - 2 - 0 = -1\,.$$

Going *against* one of the oriented edges, makes us count the flow with the opposite sign.

Recall that an oriented edge E_i of a partition in $\mathbf{R}^n$ is a vector that goes with or against the edge and any collection of such edges $C = \{E_i :\ i = 0, 1, ..., n\}$ is seen as an *oriented curve*.

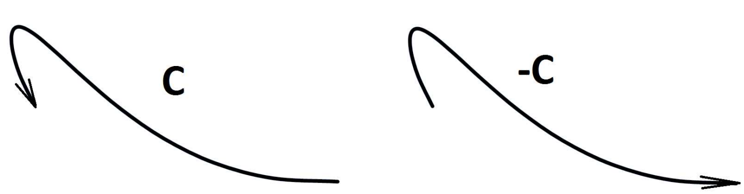
Definition 6.4.2: sum

Suppose C is an oriented curve in $\mathbf{R}^n$ that consists of oriented edges $E_i,\ i = 1, ..., n$, of a partition in $\mathbf{R}^n$. If a function G defined at the secondary nodes at the edges of the partition in $\mathbf{R}^n$ and, in particular, at the edges $\{Q_i\}$ of the curve, then *the sum of G along curve C* is defined and denoted to be the following:

$$\boxed{\sum_C G = \sum_{i=1}^n G(Q_i)}$$

When the secondary nodes aren't specified, this sum is the sum of the real-valued 1-form G.

Unlike the arc-length, the sum depends on the direction of the trip.



This dependence is however very simple: the *sign* is reversed when the direction is reversed.

Theorem 6.4.3: Negativity of Sum

$$\sum_{-C} G = -\sum_{C} G$$

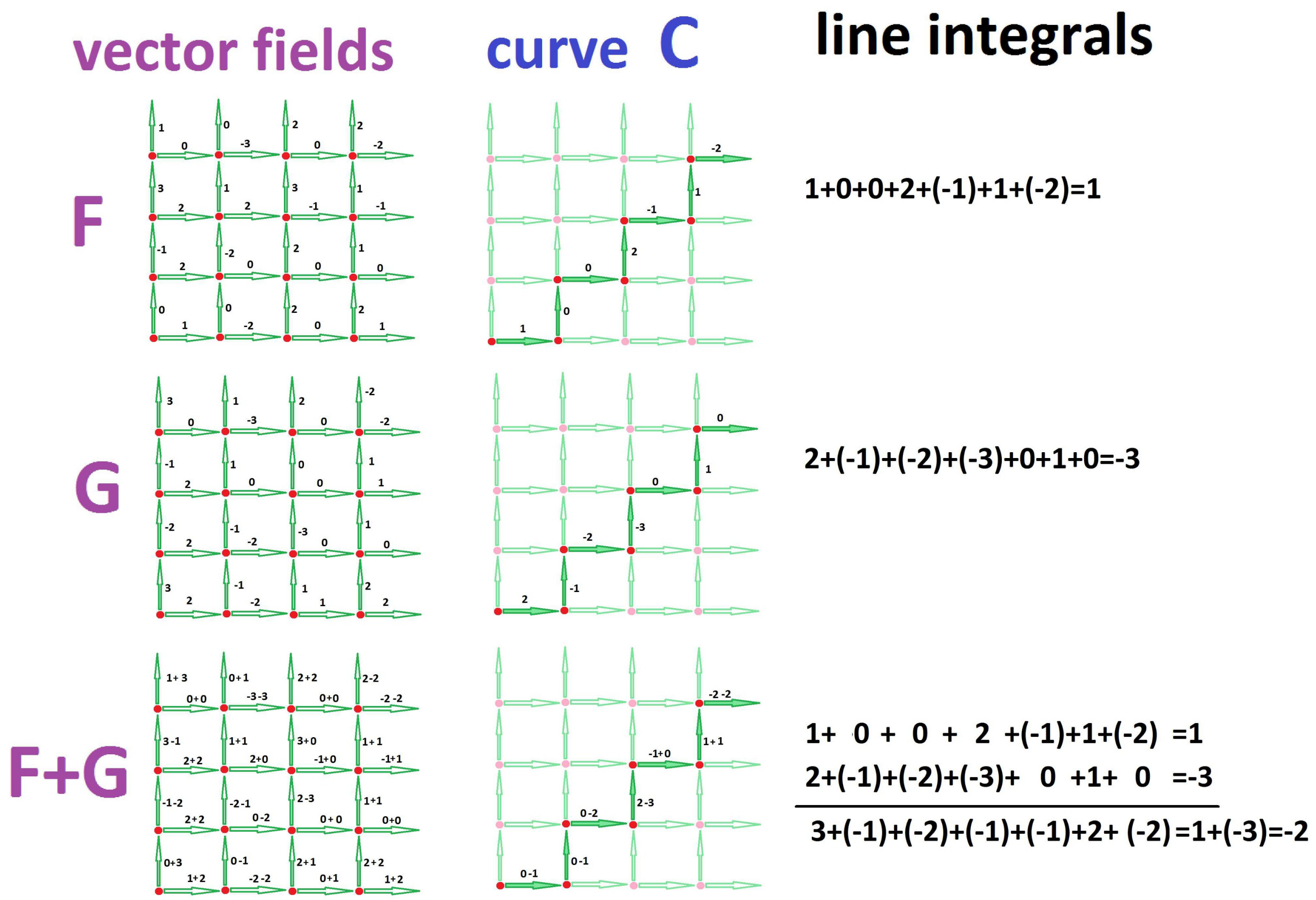
More familiar algebraic properties:

Theorem 6.4.4: Linearity of Sum

For any two functions F and G defined at the secondary nodes at the edges of the partition in $\mathbf{R}^n$ and any two numbers λ and μ, we have:

$$\boxed{\sum_{C} (\lambda F + \mu G) = \lambda \sum_{C} F + \mu \sum_{C} G}$$

The Sum Rule is illustrated below:

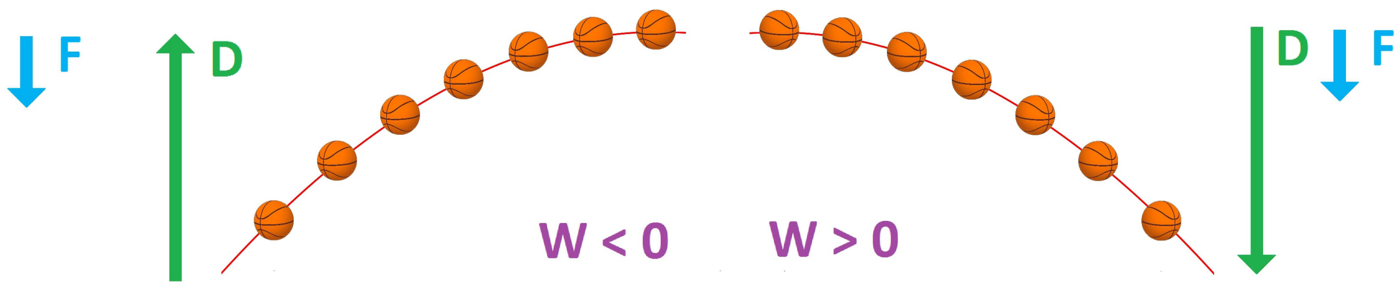


Theorem 6.4.5: Additivity of Sum

For any two oriented curves of edges C and K with no edges in common and that together form an oriented curve of edges $C \cup K$, we have:

$$\boxed{\sum_{C \cup K} F = \sum_{C} F + \sum_{K} F}$$

Let's examine another problem: the *work of a force*. Suppose a ball is thrown.



This force is directed down, just as the movement of the ball. The work done on the ball by this force as it falls is equal to the (signed) magnitude of the force, i.e., the weight of the ball, multiplied by the (signed) distance to the ground, i.e., the displacement. All horizontal motion is ignored as unrelated to the gravity. Moving an object up from the ground the work performed by the gravitational force is negative.

Of course, we are speaking of *vectors*.

In the 1-dimensional case, suppose that the force F is constant and the displacement D is along a straight line. Then the *work* W is equal to their product:

$$W = F \cdot D\,.$$

The force may vary with location, however: spring tension, gravitation, air pressure. Recall our approach:

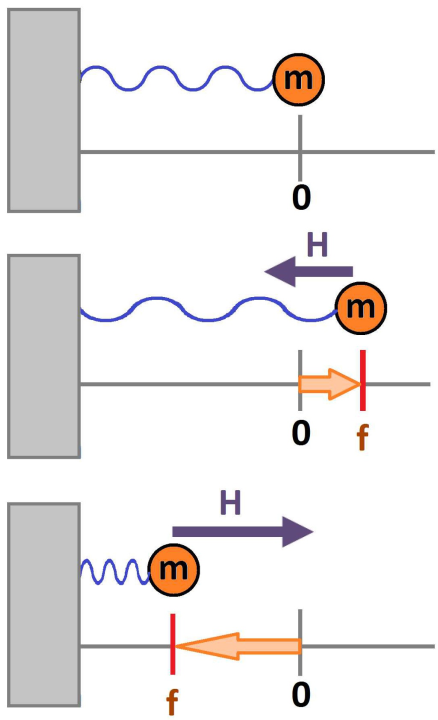
Definition 6.4.6: work

If a function F on segment $[a,b]$ is called a *force function* then its Riemann integral $\int_a^b F\,dx$ is called the *work* of the force over interval $[a,b]$.

Example 6.4.7: Hooke's Law

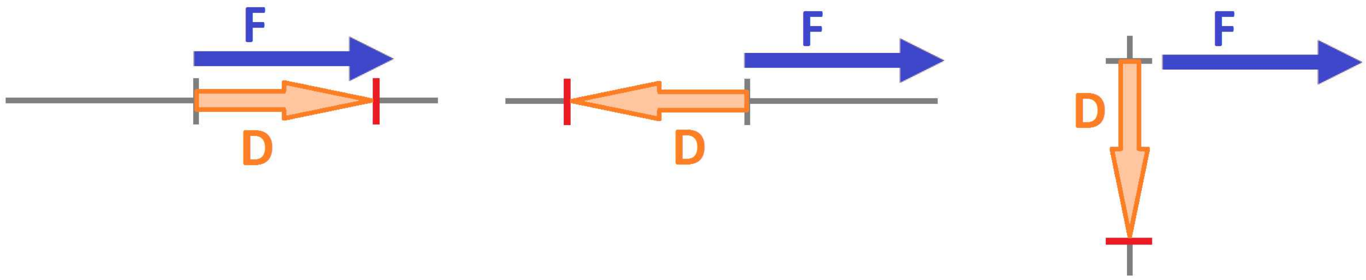
In the case of an object attached to a *spring*, the force is proportional to the (signed) distance of the object to its equilibrium:

$$F = -kx\,.$$



Let's now proceed to the n-dimensional case but start with a constant force and linear motion.

This time, the force and the displacement may be misaligned:



In addition to motion "with the force" and "against the force", the third possibility emerges: What if we move perpendicular to the force? The motion is unaffected by the force! This is the case of horizontal motion under gravity force, which is constant close to the surface of the Earth. We conclude:

▶ When the motion is perpendicular to the force, the work is zero.

Furthermore, what if the direction of our path varies but only within the standard square *grid* on the plane? We realize that there is a force vector associated with each edge of our trip and possibly with every edge of the grid. However, only one of these vector components matters: the horizontal when the edge is horizontal and the vertical when the edge is vertical. It is then sufficient to assign this *single* number to each edge to indicate the force applied to this part of the trip.

Example 6.4.8: system of pipes

As a familiar interpretation, we can look at this as a system of *pipes* with the numbers indicating the speed of the flow in each pipe (along the directions of the axes). If, for example, we are moving through a grid with $\Delta x \times \Delta y$ cells, the work along the "staircase" is the following:

$$W = 1 \cdot \Delta x + 0 \cdot \Delta y + 0 \cdot \Delta x + 2 \cdot \Delta y + (-1) \cdot \Delta x + 1 \cdot \Delta y + (-2) \cdot \Delta x \,.$$

When $\Delta x = \Delta y = 1$, this is simply the sum of the values provided:

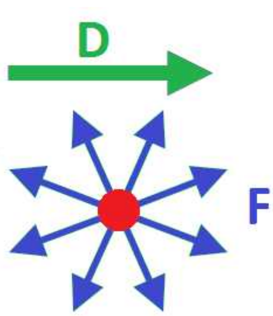
$$W = 1 + 0 + 0 + 2 + (-1) + 1 + (-2) = 1 \,.$$

What if we, instead, go around the first square? Then

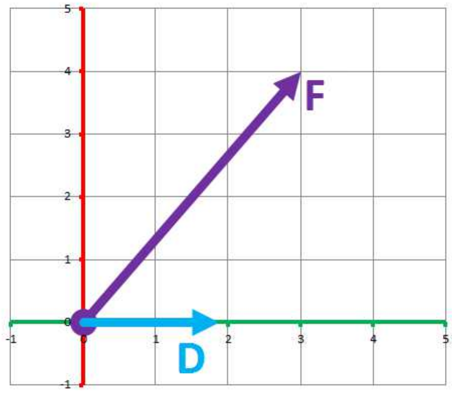
$$W = 1 + 0 - 2 - 0 = -1 \,.$$

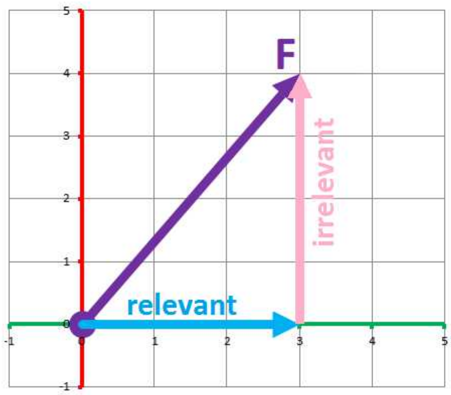
Going *against* one of the oriented edges, makes us count the work with the opposite sign. In other words, the edge and the displacement are multiples of each other.

When the direction of the force isn't limited to the grid anymore, it can take, of course, one of the diagonal directions. In fact, there is a whole circle of possible directions:



The vector of the force, too, can take all available directions. We would like to find and discard the irrelevant (perpendicular) part of the force F:





We decompose it into *parallel and normal components* relative to the displacement:

$$\boxed{F = F_{\perp} + F_{||}}$$

The relevant ("collinear") component of the force F is the projection on the displacement vector:

$$F_{||} = ||F|| \cos \alpha \,,$$

where α is the angle of F with D.

Of course, we are talking about the *dot product* here.

The *work of the force vector F along the displacement vector D* is defined to be their dot product:

$$\boxed{W = F \cdot D}$$

The work is proportional to the magnitude of the force and to the magnitude of the displacement. It is also proportional the projection of the former on the latter (the relevant part of the force) and the latter on the former (the relevant part of the displacement). It makes sense.

In our interpretation of a vector field as a system of *pipes* has two numbers, this is a vector associated with each pipe indicating the speed of the flow in the pipe (along the direction of one of the pipe) as well as the leakage (perpendicular to this direction). Then, the relevant part of the force is found as the (scalar) *projection* of the vector of the force on the vector of displacement. The difference is between real-valued and vector-valued 1-forms.

Thus, the work is represented as the dot product of the vector of the force and the vector of displacement.

Now the work over a whole trip:

Definition 6.4.9: Riemann sum

Suppose C is an oriented curve in $\mathbf{R}^n$ that consists of oriented edges E_i, $i = 1, ..., n$, of a partition in $\mathbf{R}^n$. If a vector field F is defined at the secondary nodes at the edges of the partition in $\mathbf{R}^n$ and, in particular, at the edges $\{Q_i\}$ of the curve, then *the Riemann sum of F along curve C* is defined and denoted to be the following:

$$\boxed{\sum_C F \cdot \Delta X = \sum_{i=1}^{n} F(Q_i) \cdot E_i}$$

In other words, this is the Riemann sum of a vector field, F, is the sum of a certain real-valued function, $F \cdot E$, along a curve as defined in the beginning of the section.

Definition 6.4.10: work

When the vector field F is called a *force field*, then the sum of F along C is also called the *work of force F along curve C*.

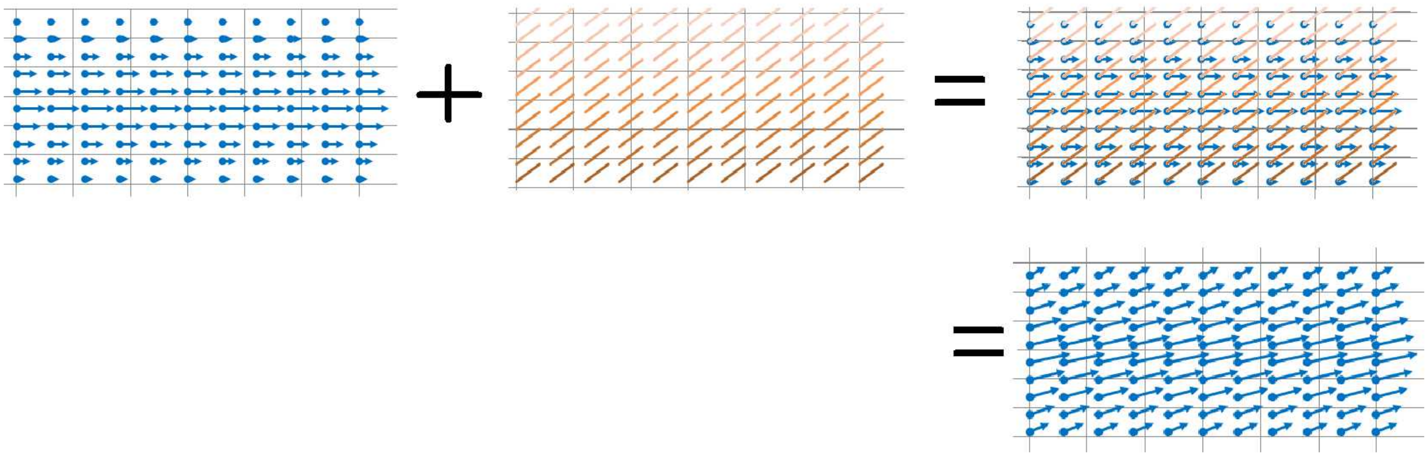
Warning!

Only the part of the force field passed through by the object affects the work.

The properties follow the ones presented above:

Theorem 6.4.11: Negativity of Riemann Sum

$$\sum_{-C} F \cdot \Delta X = -\sum_{C} F \cdot \Delta X$$



Theorem 6.4.12: Linearity of Riemann Sum

For any two vector fields F and G defined at the secondary nodes at the edges of the partition in $\mathbf{R}^n$ and any two numbers λ and μ, we have:

$$\boxed{\sum_{C} (\lambda F + \mu G) \cdot \Delta X = \lambda \sum_{C} F \cdot \Delta X + \mu \sum_{C} G \cdot \Delta X}$$

Theorem 6.4.13: Additivity of Riemann Sum

For any two oriented curves C and K with only finitely many points in common and that together form an oriented curve $C \cup K$, we have:

$$\boxed{\sum_{C \cup K} F \cdot \Delta X = \sum_{C} F \cdot \Delta X + \sum_{K} F \cdot \Delta X}$$

6.5. Line integrals: work

A more general setting is that of a motion through space, $\mathbf{R}^n$, with a *continuously changing force.*

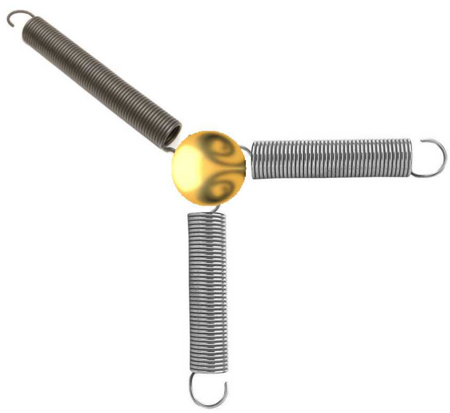
We first assume that we move from point to point along a *straight line.*

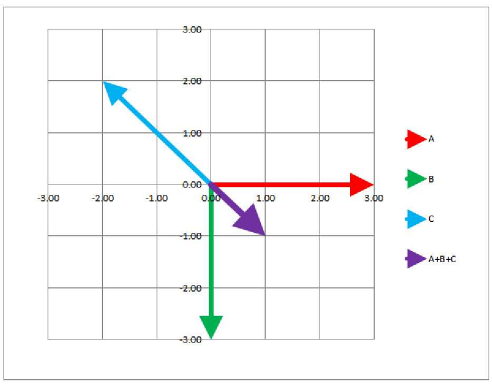
Example 6.5.1: Newton's Law of Gravity

Away from the ground, the *gravity* is proportional to the reciprocal of the square of the distance of the object to the center of the planet:

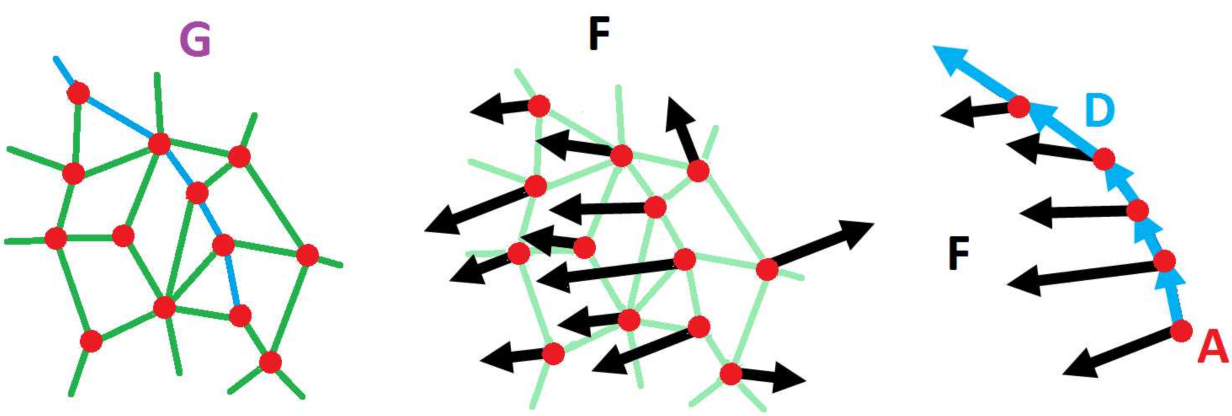
$$F(X) = -\frac{kX}{||X||^3}\,.$$

The *pressure* and, therefore, the medium's resistance to motion may change arbitrarily. Multiple springs create a 2-dimensional variability of forces:





The definition of work applies to straight travel... or to travel along multiple straight edges:



If these segments are given by the displacement vectors $D_1, ..., D_n$ and the force for each is given by the vectors $F_1, ..., F_n$, then the work is defined to be the simple sum of the work along each:

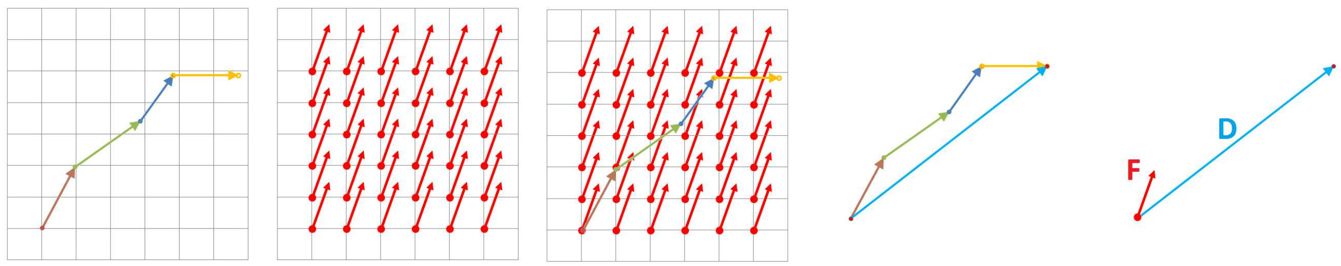
$$W = F_1 \cdot D_1 + ... + F_n \cdot D_n\,.$$

Example 6.5.2: constant force

If the force is constant $F_i = F$, we simplify,

$$W = F \cdot D_1 + ... + F \cdot D_n = F \cdot (D_1 + ... + D_n)\,,$$

and discover that the total work is the dot product of the force and the *total displacement.*



This makes sense. This is a simple example of "path-independence".

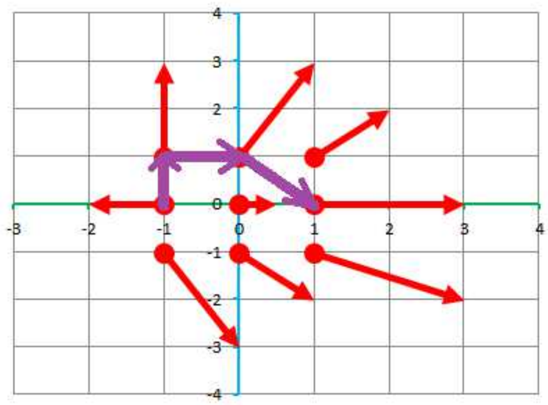
Furthermore, the round trip will require zero work... unless one has to walk to school "5 miles – uphill both ways!" The issue isn't as simple as it seems: even though it is impossible to make round trip while walking uphill, it is possible during this trip to walk against the wind even though the wind doesn't change. It all depends on the nature of the vector field.

Example 6.5.3: adding work

In order to compute the work of a vector field along a curve made of straight edges, all we need is the formula:

$$W = F_1 \cdot D_1 + ... + F_n \cdot D_n \, .$$

In order for the computation to make sense, the edges of the path and the vectors of the force have to be paired up! Here's a simple example:



We pick the value of the force from the *initial* point of each edge:

$$W =< -1, 0 > \cdot < 0, 1 > + < 0, 2 > \cdot < 1, 0 > + < 1, 2 > \cdot < 1, 1 >= 3 \, .$$

Example 6.5.4: data flow

It is possible that there is *no vector field* and the force is determined entirely by our motion. For example, the air or water resistance is directed against our velocity (and is proportional to the speed).

The above computations remain the same.

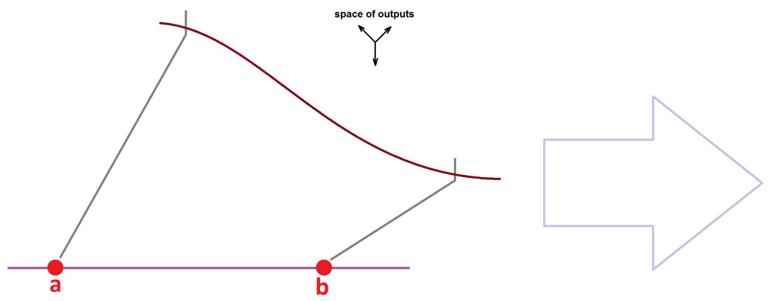
The general setup for defining and computing work along a curve is identical to what we have done several times.

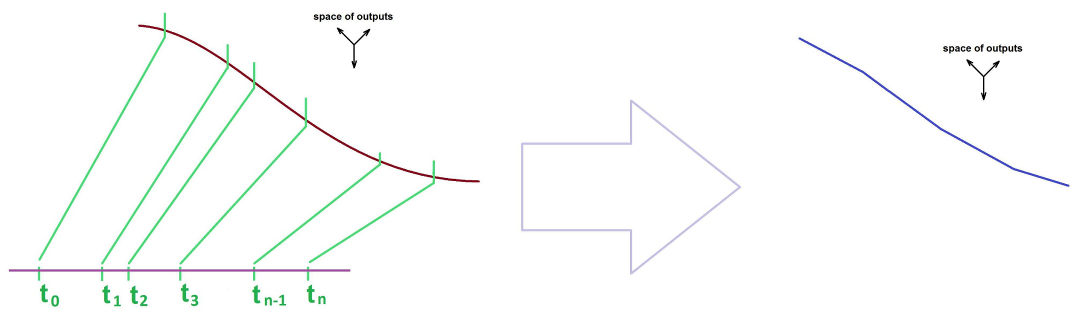
Suppose we have a sequence of points $P_i,\ i=0,1,...,n$, in $\mathbf{R}^n$. We will treat this sequence as an oriented curve C by representing it as the path of a *parametric curve* as follows. Suppose we have a sampled partition of an interval $[a,b]$:

$$a=t_0\le c_1\le t_1\le ...\le c_n\le t_n=b\,.$$

We define a parametric curve by:

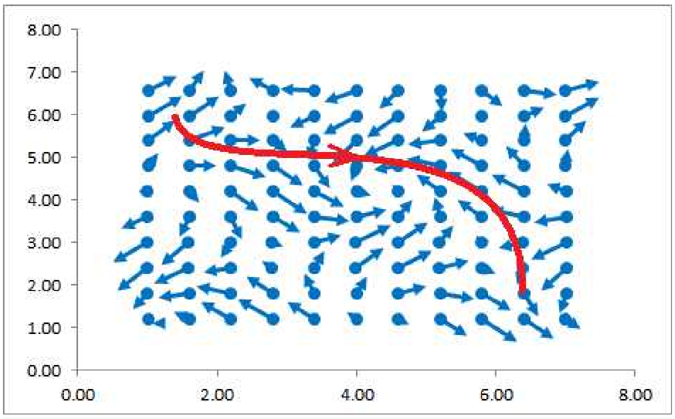
$$X(t_i)=P_i,\ i=0,1,...,n\,.$$





However, it doesn't matter how fast we go along this path. It is the path itself – the locations we visit – that matters. The direction of the trip matters too. This is then about an *oriented curve*. In the meantime, a non-constant vectors along the path typically come from a *vector field*, $F=F(X)$. If its vectors change incrementally, one may be able to compute the work by a simple summation, as above.

We then find a regular parametrization of the latter: a parametric curve $X=X(t)$ defined on the interval $[a,b]$. We divide the path into small segments with end-points $X_i=X(t_i)$ and then sample the force at the points $Q_i=X(c_i)$.



Then the work along each of these segments is approximated by the work with the force being constantly equal to $F(Q_i)$:

$$\text{work along } i\text{th segment} \approx \text{ force } \cdot \text{ length} = F(Q_i)\cdot \Delta X_i\,,$$

where ΔX_i is the displacement along the ith segment. Then,

$$\text{Total work } \approx \sum_{i=1}^{n} F(Q_i)\cdot(X_{i+1}-X_i)=\sum_{i=1}^{n} F(X(c_i))\cdot(X(t_{i+1})-X(t_i))\,.$$

This is the formula that we have used and will continue to use for approximations. Note that this is just the sum of a discrete 1-form.

Example 6.5.5: estimating

Let's estimate the work of the force field

$$F(x,y) =< xy,\ x-y >$$

along the upper half of the unit circle directed counterclockwise. First, we parametrize the curve:

$$X(t) =< \cos t,\ \sin t >,\ 0 \le t \le \pi\,.$$

We choose $n = 4$ intervals of equal length with the left-ends as the secondary nodes:

$x_0 = 0$	$x_1 = \pi/4$	$x_2 = \pi/2$	$x_3 = 3\pi/4$	$x_4 = \pi$
$c_1 = 0$	$c_2 = \pi/4$	$c_3 = \pi/2$	$c_4 = 3\pi/4$	
$X_0 = (1,\ 0)$	$X_1 = (\sqrt{2}/2,\ \sqrt{2}/2)$	$X_2 = (0,\ 1)$	$X_3 = (-\sqrt{2}/2,\ \sqrt{2}/2)$	$X_4 = (-1,\ 0)$
$Q_1 = (1,\ 0)$	$Q_2 = (\sqrt{2}/2,\ \sqrt{2}/2)$	$Q_3 = (0,\ 1)$	$Q_4 = (-\sqrt{2}/2,\ \sqrt{2}/2)$	
$F(Q_1) =< 0,\ 1 >$	$F(Q_2) =< 1/2,\ 0 >$	$F(Q_3) =< 0,\ -1 >$	$F(Q_4) =< -1/2,\ -\sqrt{2} >$	

Then,

$$\begin{aligned} W\ &\approx < 0,1 > \cdot < \sqrt{2}/2-1,\ \sqrt{2}/2 > + < 1/2,0 > \cdot < -\sqrt{2}/2,\ 1-\sqrt{2}/2 > \\ &+ < 0,-1 > \cdot < -\sqrt{2}/2,\ \sqrt{2}/2-1 > + < -1/2,\ -\sqrt{2} > \cdot < -1+\sqrt{2}/2,\ -\sqrt{2}/2 > \\ &= ... \end{aligned}$$

To bring the full power of the calculus machinery, we, once again, proceed to convert the expression into the Riemann sum of a certain function over this partition:

$$\text{total work } \approx \sum_{i=1}^{n} F(X(c_i)) \cdot \frac{X(t_{i+1})-X(t_i)}{t_{i+1}-t_i}(t_{i+1}-t_i) = \sum_{a}^{b} \left((F \circ X) \cdot \frac{\Delta X}{\Delta t} \right) \Delta t\,.$$

Then, we define the work of the force as the limit, if it exists, of these Riemann sums, i.e., the Riemann integral.

Definition 6.5.6: line integral

Suppose C is an oriented curve in $\mathbf{R}^n$. For a vector field F in $\mathbf{R}^n$, the *line integral of F along C* is denoted and defined to be the following:

$$\boxed{\int_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt}$$

where $X = X(t),\ a \le t \le b$, is a regular parametrization of C.

Definition 6.5.7: work

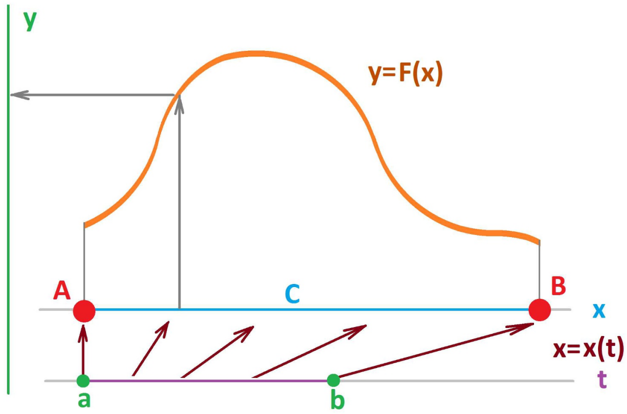
When the vector field F is called a *force field*, then the integral of F along C is also called the *work of force F along curve C*.

The first term in the integral shows how the force varies with time during our trip. Just as always, the

Leibniz notation reveals the meaning:

$$\int_C F \cdot dX = \int_a^b (F \circ X) \cdot \frac{dX}{dt} dt \,,$$

Once all the vector algebra is done, we are left with just a familiar numerical integral from Volume 2 (Chapter 2DC-6). Furthermore, when $n = 1$, the integral *is* the familiar numerical integral from Chapter 2DC-6. Indeed, suppose $x = F(t)$ is just a numerical function and C is the interval $[A, B]$ in the x-axis.



Then we have:

$$\int_C F \cdot dX = \int_{x=A}^{x=B} F \, dx = \int_{t=a}^{t=b} F(x(t))x'(t) \, dt \,,$$

where $x = x(t)$ serves as a parametrization of this interval so that $x(a) = A$ and $x(b) = B$. This is just an interpretation of the *integration by substitution* formula.

Example 6.5.8: straight path

Compute the work of a constant vector field, $F =< -1, 2 >$, along a straight line, the segment from $(0, 0)$ to $(1, 3)$. First, parametrize the curve and find its derivative:

$$X(t) =< 1, 3 > t, \ 0 \le t \le 1, \implies X'(t) =< 1, 3 > \,.$$

Then,

$$W = \int_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t) \, dt = \int_0^1 < -1, 2 > \cdot < 1, 3 > \, dt = \int_0^1 5 \, dt = 5 \,.$$

Example 6.5.9: radial vector field

Compute the work of the *radial vector field*, $F(X) = X =< x, y >$, along the upper *half-circle* from $(1, 0)$ to $(-1, 0)$. First parametrize the curve and find its derivative:

$$X(t) =< \cos t, \ \sin t >, \ 0 \le t \le \pi, \implies X'(t) =< -\sin t, \cos t > \,.$$

Then,

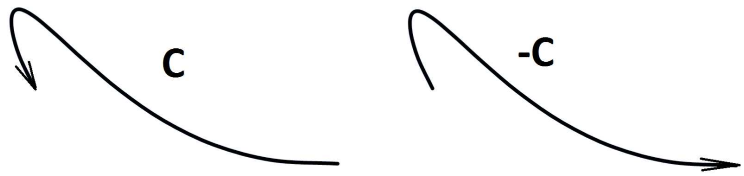
$$\begin{aligned} W &= \int_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t) \, dt \\ &= \int_0^\pi < \cos t, \ \sin t > \cdot < -\sin t, \ \cos t > \, dt \\ &= \int_0^\pi (\cos t(-\sin t) + \sin t \cos t) \, dt \\ &= 0 \,. \end{aligned}$$

Theorem 6.5.10: Independence of Work

The work is independent of parametrization.

Thus, just as we used parametric curves to study a function of several variables, we use them to study a vector field. Note however, that only the part of the vector field visited by the parametric curve affects the line integral.

Unlike the arc-length, the work depends on the direction of the trip.



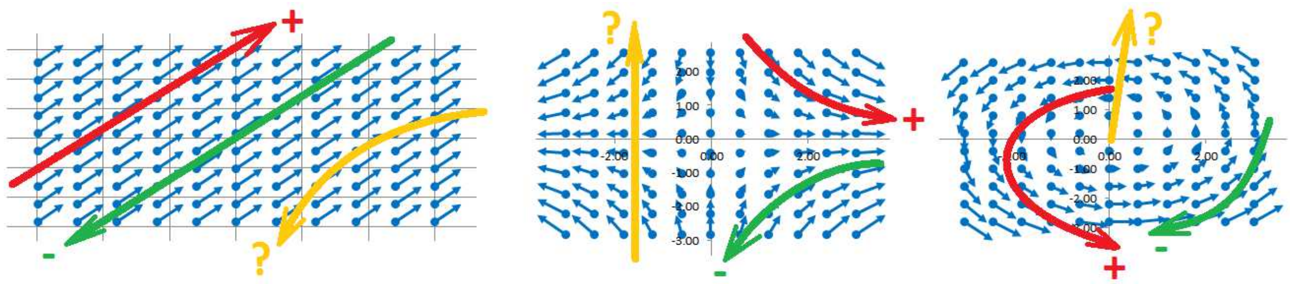
This dependence is however very simple: the *sign* is reversed when the direction is reversed.

Theorem 6.5.11: Negativity of Integral

$$\int_{-C} F \cdot dX = -\int_C F \cdot dX$$

Example 6.5.12: qualitative analysis

Is the work positive or negative?



When all the angles are acute, it's positive.

Exercise 6.5.13

Finish the example.

Exercise 6.5.14

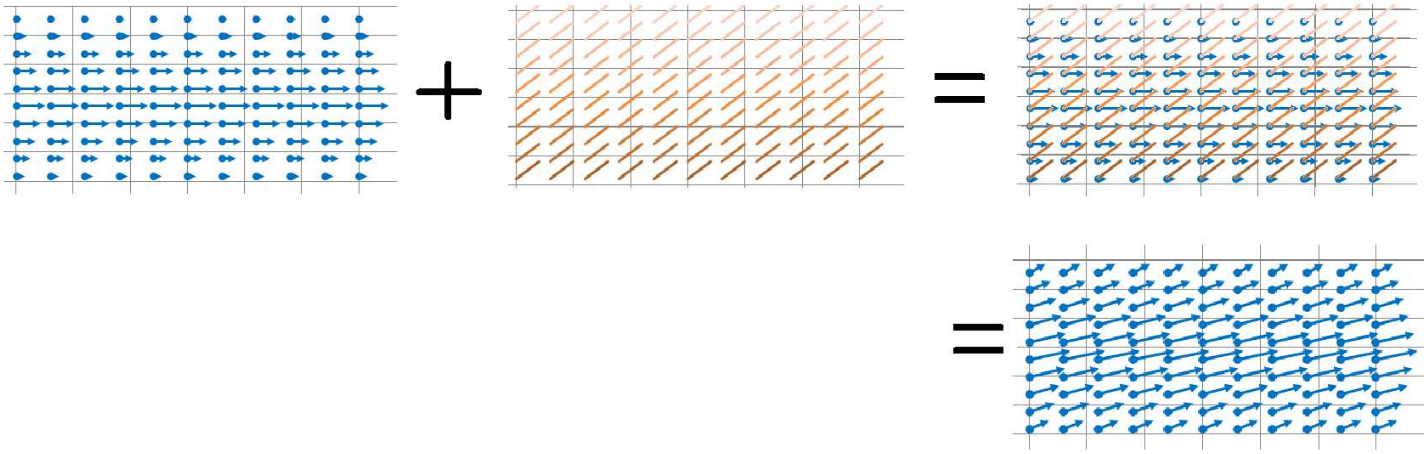
How much work does it take to move an object attached to a spring s units from the equilibrium?

Exercise 6.5.15

How much work does it take to move an object s units from the center of a planet?

Exercise 6.5.16

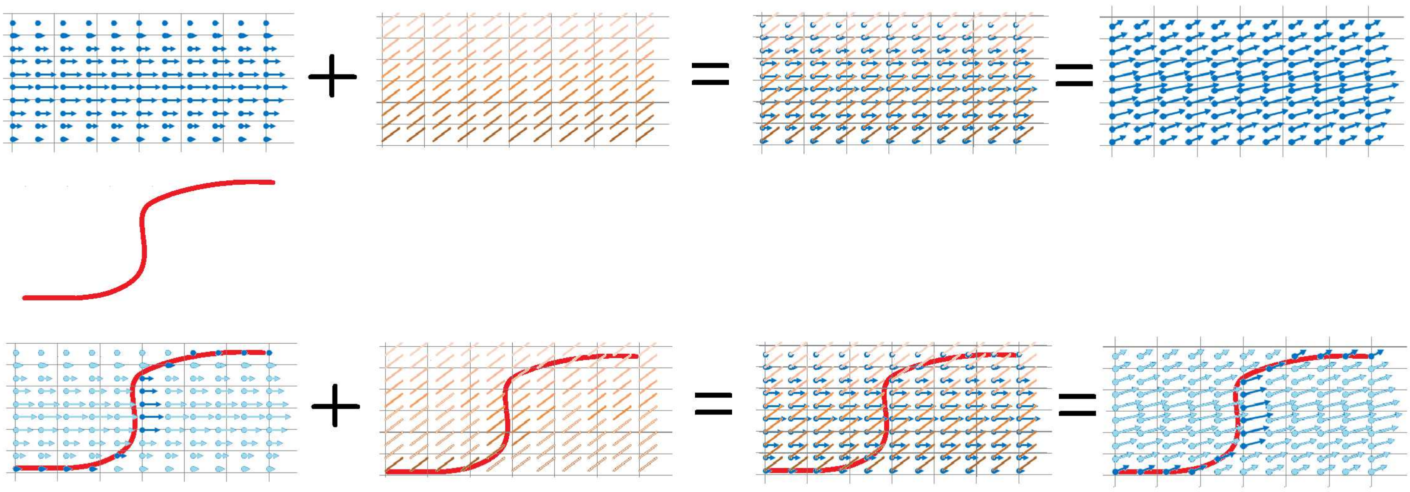
What is the value of the line integral of the gradient of a function along one of its level curves?



Theorem 6.5.17: Linearity of Integral

For any two vector fields F and G and any two numbers λ and μ, we have:

$$\boxed{\int_C (\lambda F + \mu G) \cdot dX = \lambda \int_C F \cdot dX + \mu \int_C G \cdot dX}$$



Theorem 6.5.18: Additivity of Integral

For any two oriented curves C and K with only finitely many points in common and that together form an oriented curve $C \cup K$, we have:

$$\boxed{\int_{C\cup K} F \cdot dX = \int_C F \cdot dX + \int_K F \cdot dX}$$

Let's look at the *component representation of the integral.* Starting with dimension $n = 1$, the definition,

$$\int_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt\,,$$

becomes ($F = f,\ X = x,\ C = [A, B]$):

$$\int_A^B f(x)\, dx = \int_a^b f(x(t))x'(t)\, dt\,,$$

where $A = x(a)$ and $B = x(b)$. In $\mathbf{R}^2$, we have the following component representation of a vector field F and the increment of X:

$$F =< p, q > \quad \text{and} \quad dX =< dx, dy > \,.$$

Then the line integral of F along C is denoted as follows:

$$\int_C F\cdot dX = \int_C < p,q > \cdot < dx,dy >= \int_C p\,dx + q\,dy\,.$$

Here, the integrand is a *differential form of degree* 1:

$$p\,dx + q\,dy$$

The notation matches the formula of the definition. Indeed, the curve's parametrization $X = X(t),\ a \le t \le b$, has a component representation:

$$X =< x,y >\,,$$

therefore,

$$\int_a^b F(X(t))\cdot X'(t)\,dt = \int_a^b F(x(t),y(t))\cdot < x'(t),y'(t) > \,dt = \int_a^b p(x(t),y(t))x'(t)\,dt + \int_a^b q(x(t),y(t))y'(t)\,dt\,.$$

Similarly, in $\mathbf{R}^3$, we have a component representation of a vector field F and the increment of X:

$$F =< p,q,r > \quad \text{and} \quad dX =< dx,dy,dz > \ .$$

Then the line integral of F along C is denoted as follows:

$$\int_C F\cdot dX = \int_C p\,dx + q\,dy + r\,dz\,.$$

Let's review the recent integrals that involve parametric curves. Suppose $X = X(t)$ is a parametric curve on $[a,b]$.

• The first is the (componentwise) *integral of the parametric curve*:

$$\int_a^b X(t)\,dt\,,$$

providing the displacement from the known velocity, as functions of time.

• The second is the *arc-length integral*:

$$\int_C f\,ds = \int_a^b f(X(t))||X'(t)||\,dt\,,$$

providing the mass of a curve of variable density.

• The third is the *line integral along an oriented curve*:

$$\int_C F\cdot dX = \int_a^b F(X(t))\cdot X'(t)\,dt\,,$$

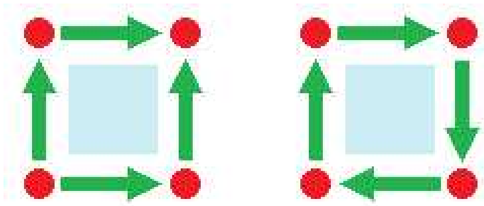
providing the work of the force field.

The main difference between the first and the other two is that in the former case the parametric curve is the *integrand* (and the output is another parametric curve) and in the latter it provides the *domain of integration* (and the output is a number).

6.6. Sums along closed curves reveal exactness

Example 6.6.1: single cell

Let's consider the curve C that goes around a single square of the grid counterclockwise. Let G be a *constant* function (1-form) on the partition: it has same value for each horizontal edge and same for each vertical edge (left):



Then the flow along the curve is *zero*! Note that G is exact:

$$G = \Delta f\,.$$

These are G (defined on edges) and f (defined on nodes):

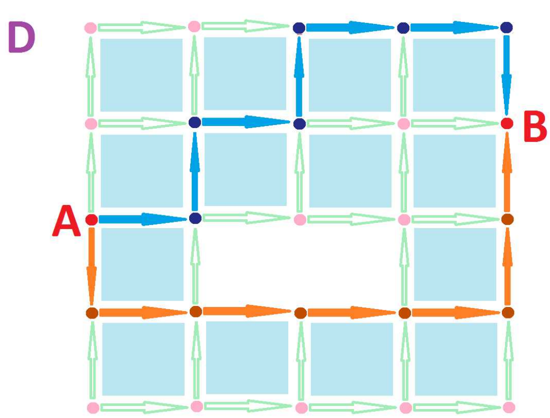
$$G = \boxed{\begin{array}{ccc} \bullet & 1 & \bullet \\ 1 & & 1 \\ \bullet & 1 & \bullet \end{array}} \quad f = \boxed{\begin{array}{ccc} 2 & -- & 3 \\ | & & | \\ 1 & -- & 2 \end{array}}$$

Suppose now that G is *rotational* (right):

$$G = \boxed{\begin{array}{ccc} \bullet & 1 & \bullet \\ 1 & & -1 \\ \bullet & -1 & \bullet \end{array}} \quad f = ?$$

Then the flow is *not zero*! Note that G isn't exact, as demonstrated in the first section of the chapter. This isn't a coincidence.

Suppose C is an oriented curve that consists of oriented edges $Q_i,\ i = 1, ..., m$, of a partition of a region D in $\mathbf{R}^n$.



Suppose a function defined on the secondary nodes F is the difference in D, $G = \Delta f$, of some function f defined on the primary nodes of the partition. We carry out a familiar computation by adding all of these

and canceling the repeated nodes:

$$\begin{array}{llllll}\sum_C G & = G(Q_1) & +G(Q_2) & +\dots & +G(Q_m) \\ & = G(P_0P_1) & +G(P_1P_2) & +\dots & +G(P_{m-1}P_m) \\ & = \big[f(P_1)-f(P_0)\big] & +\big[f(P_2)-f(P_1)\big] & +\dots & +\big[f(P_m)-f(P_{m-1})\big] \\ & = -f(P_0) & & & +f(P_m) \\ & = f(B)-f(A)\,. \end{array}$$

We have proven the following.

Theorem 6.6.2: Fundamental Theorem of Calculus for Differences II

Suppose a function defined on the secondary nodes G is exact, i.e., $G = \Delta f$ for some function f defined on the primary nodes of the partition of region D. If an oriented curve C in D starts at node A and ends at node B, then we have:

$$\boxed{\sum_C G = f(B) - f(A)}$$

Now, the sum on right is independent of our choice of C as long as it is from A to B! We formalize this property below.

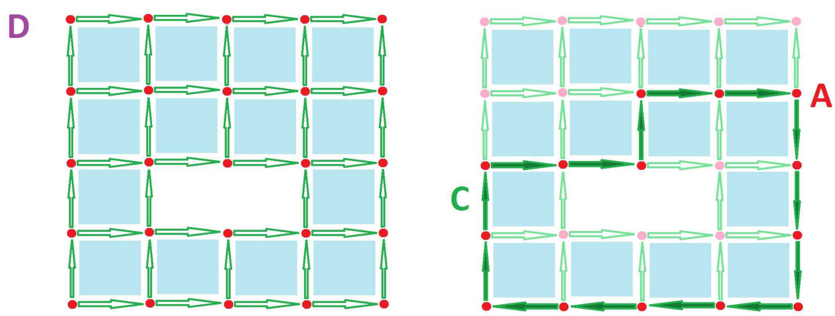
Definition 6.6.3: path-independent function

A function defined on the secondary nodes of a partition of a region D in $\mathbf{R}^n$ is called *path-independent over D* if its sum along any oriented curve depends only on the start- and the end-points of the curve; i.e.,

$$\boxed{\sum_C G = \sum_K G}$$

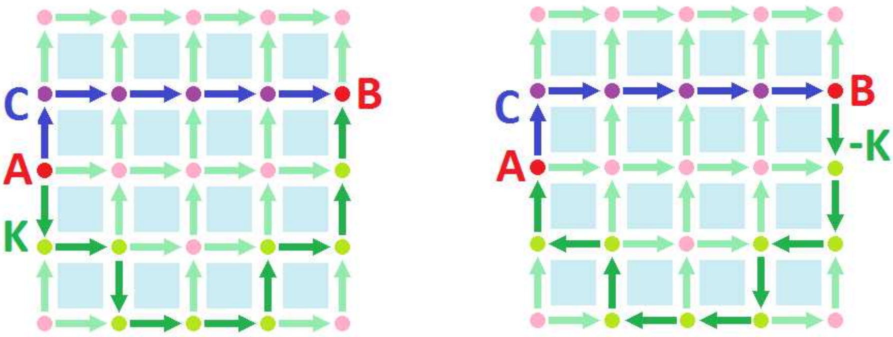
for any two curves of edges C and K from node A to node B that lie entirely in D.

What can we say about the sums of such functions along *closed* curves?



The path-independence allows us to compare the curve to any curve with the same end-points.

What is the simplest one to compare to? Consider this: If there are no pipes, there is no flow! We are talking about a special kind of path, a *constant curve*: $K = \{A\}$. Let's compare it to another curve C:



The curve K is trivial; therefore, we have:

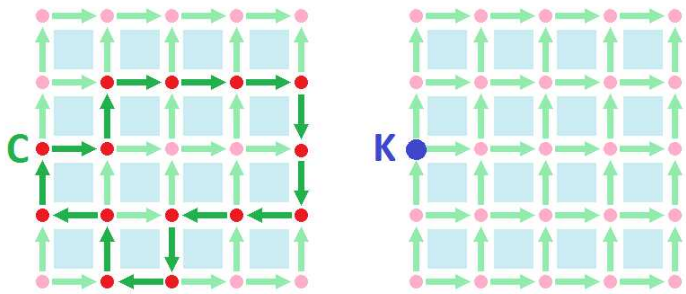
$$\sum_C G = \sum_K G = 0\,.$$

So, path-independence implies zero sums along any closed curve.

The converse is also true. Suppose we have two curves C and K from A to B. We create a new, *closed* curve from them. We glue C and the reversed K together:

$$Q = C \cup -K\,.$$

It goes from A to A.



Then, from *Additivity* and *Negativity* we have:

$$0 = \sum_Q G = \sum_C G + \sum_{-K} G = \sum_C G - \sum_K G\,.$$

Therefore,

$$\sum_C G = \sum_K G\,.$$

In summary, we have the following:

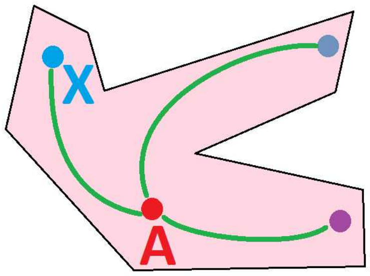
Theorem 6.6.4: Path-independence

A function defined on the secondary nodes of a partition of a region D in $\mathbf{R}^n$ is path-independent if and only all of its sums along closed curves in the partition are equal to zero.

Suppose we have a path-independent function G defined on edges of a partition of some set D in $\mathbf{R}$. We know it to be exact, but how do we find f with $\Delta f = G$? The idea comes from Volume 3 (Chapter 3IC-1). First, we choose an arbitrary node A in D and then carry out a summation along *every* possible curve from A. We define for each X in D:

$$f(X) = \sum_C G\,,$$

where C is any curve from A to X. A choice of C doesn't matter because G is path-independent.



To ensure that this function is well defined we need an extra requirement.

Theorem 6.6.5: Fundamental Theorem of Calculus for Differences I

On a partition of a path-connected region D in $\mathbf{R}^n$, if $G = \Delta f$, the function below is well-defined for a fixed A in D:

$$g(X) = \sum_C G\,,$$

where C is any curve from A to X within the partition of D, and, furthermore,

$$\boxed{\Delta g = G}$$

Proof.

Because the region in path-connected, there is always a curve from A to any X.

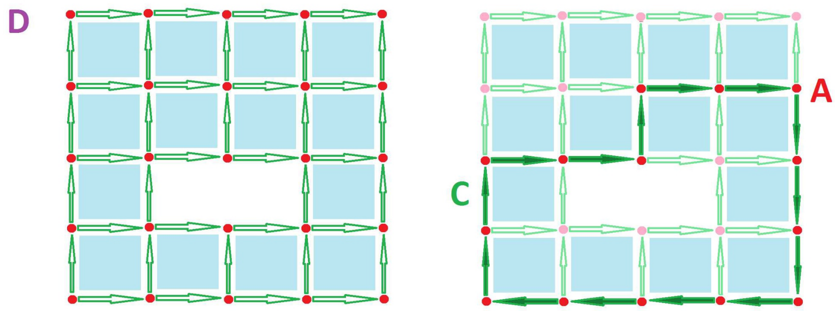
What about vector fields? If F is a vector field, we apply the above analysis to its projection $G = F \cdot \Delta X$. The sums become Riemann sums...

Example 6.6.6: constant vector field

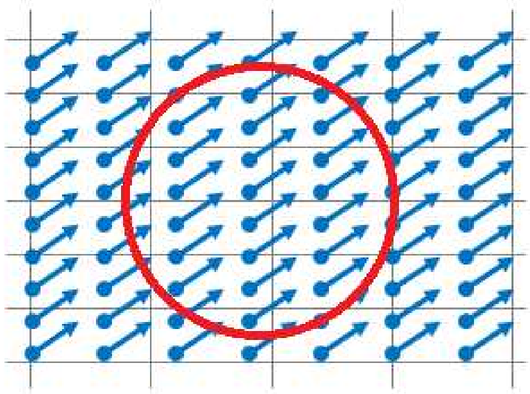
Suppose C is an oriented curve that consists of oriented edges Q_i, $i = 1, ..., m$, of a partition of a region D in $\mathbf{R}^n$. Let

$$Q_i = P_{i-1}P_i \text{ with } P_0 = P_m = A\,.$$

It may look like this:



Suppose F is constant vector field in D; i.e., $F(X) = G$ for all X in D:



Then the work of G along C is the following Riemann sum:

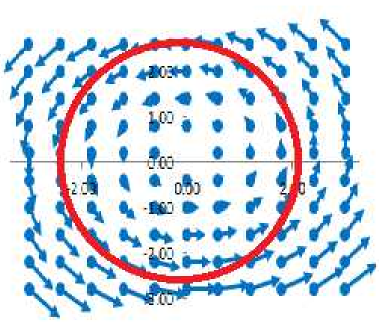
$$\begin{aligned}\sum_C F\cdot\Delta X &= \sum_{i=1}^m F(Q_i)\cdot Q_i\\ &= \sum_{i=1}^m F\cdot Q_i\\ &= F\cdot\sum_{i=1}^m P_{i-1}P_i\\ &= F\cdot\sum_{i=1}^m (P_i-P_{i-1})\\ &= F\cdot\big[(P_1-P_0)+(P_2-P_1)+...+(P_m-P_{m-1})\big]\\ &= F\cdot\big[-P_0+P_m\big]\\ &= 0\,.\end{aligned}$$

The work is *zero*!

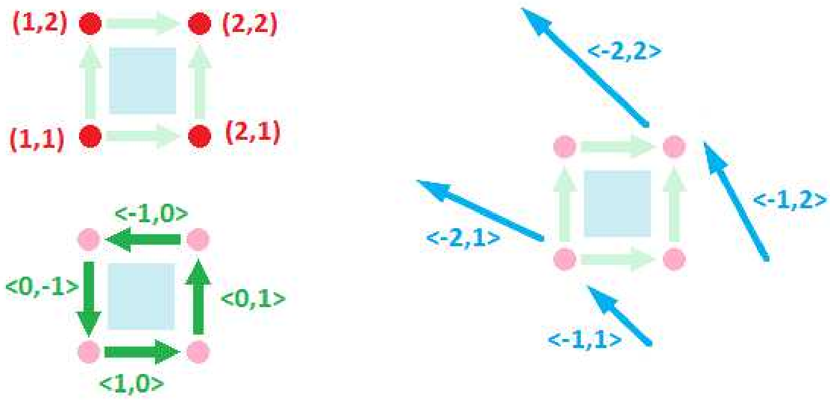
Example 6.6.7: rotation vector field

The story is the exact opposite for the *rotation vector field*:

$$F=<-y,x>\ .$$



Let's consider a single square of the partition; for example, $S=[1,2]\times[1,2]$.



Suppose curve C goes counterclockwise and the secondary nodes are the starting points of the edges. Then the work of G along C is the following Riemann sum:

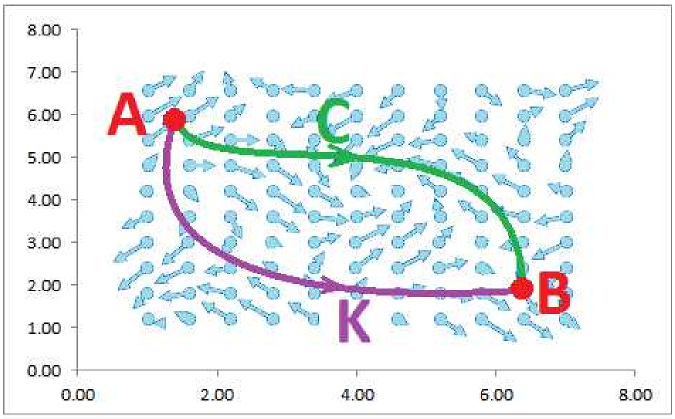
$$\begin{aligned}
&\sum_C F\cdot\Delta X =\\
&=\sum_{i=1}^{4} F(Q_i)\cdot Q_i\\
&=F(1,1)\cdot<1,0>+F(2,1)\cdot<0,1>+F(2,2)\cdot<-1,0>+F(1,2)\cdot<0,-1>\\
&=<-1,1>\cdot<1,0>+<-1,2>\cdot<0,1>+<-2,2>\cdot<-1,0>+<-2,1>\cdot<0,-1>\\
&=-1+2+2-1\\
&=2\,.
\end{aligned}$$

The work is *not zero*!

The above formula for differences takes the following form:

$$\sum_C F\cdot\Delta X = f(B)-f(A)\,.$$

Not only the proof but also the formula itself looks like the familiar Fundamental Theorem of Calculus for numerical integrals from Volume 3 (Chapter 3IC-1).



Definition 6.6.8: path-independent vector field

A vector field F defined on the secondary nodes of a partition of a region D in $\mathbf{R}^n$ is called *path-independent* if its projection $F\cdot\Delta X$ is; i.e., the Riemann sum along any oriented curve depends only on the start- and the end-points of the

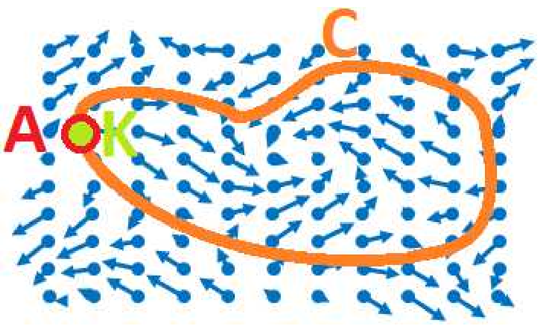
curve:

$$\boxed{\sum_C F \cdot \Delta X = \sum_K F \cdot \Delta X}$$

for any two curves of edges C and K from node A to node B that lie entirely in D.

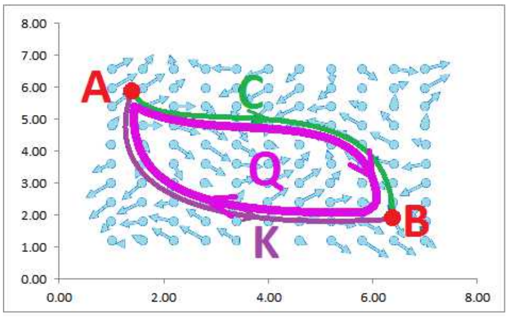
For the sum along a *closed* curve, we note once again: if we stay home, we don't do any work! We have for a path-independent vector field F:

$$\sum_C F \cdot \Delta X = \sum_K F \cdot \Delta X = 0 .$$



Conversely, suppose we have two curves C and K from A to B. We create a new, *closed* curve from them, from A to A, by gluing C and the reversed K together:

$$Q = C \cup -K .$$



From the corresponding result for differences we derive the following:

Theorem 6.6.9: Path-independence of Vector Fields

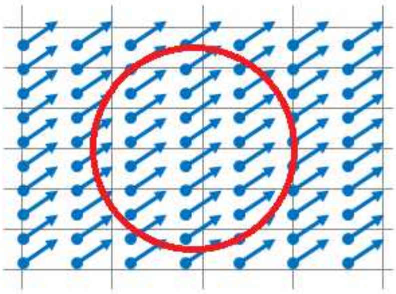
A vector field defined on the secondary nodes of a partition of a region D in $\mathbf{R}^n$ is path-independent if and only if all of its Riemann sums along closed curves of edges in D are equal to zero.

6.7. Path-independence of integrals

We will consider force fields along closed curves.

Definition 6.7.1: closed curve

A *closed curve* (a loop) is a curve that the initial and the end points of which coincidence; i.e., it is parametrized by some $X = X(t),\ a \le t \le b$, with $X(a) = X(b) = A$.



Line integrals along closed curves have a special notation:

Loop integral

$$\oint_C F \cdot dX$$

let's consider *constant* force fields along

Example 6.7.2: loop in constant vector field

Once again, what is the work of a constant force field along a closed curve such as a circle?

Consider two diametrically opposite points on the circle. The directions of the tangents to the curve are opposite while the vector field is the same. Therefore, the terms $F \cdot X'$ in the *work* integral are negative of each other. So, because of this symmetry, two opposite halves of the circle will have work negative of each other and cancel. The work must be *zero*!

Let's confirm this for $F =< p, q >$ and the standard parametrization of the circle:

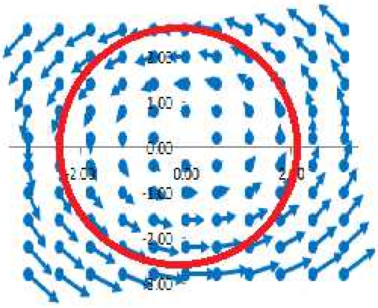
$$\begin{aligned} W &= \oint_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt \\ &= \int_0^{2\pi} < p, q > \cdot < \cos t,\ \sin t >'\ dt \\ &= \int_0^{2\pi} < p, q > \cdot < -\sin t,\ \cos t >\ dt \\ &= \int_0^{2\pi} (-p \sin t + q \cos t)\, dt \\ &= (p \cos t - q \sin t)\Big|_0^{2\pi} + (p \cos t - q \sin t)\Big|_0^{2\pi} \\ &= 0 + 0 = 0\,. \end{aligned}$$

So, work cancels out during this round trip.

Example 6.7.3: loop in rotation vector field

The story is the exact opposite for the *rotation vector field*:

$$F =< -y, x > \ .$$



Consider any point. The direction of the tangent to the curve is the same as the vector field. Therefore, the terms $F \cdot X'$ cannot cancel. The work is *not* zero! Let's confirm this result:

$$\begin{aligned}
W &= \oint_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt \\
&= \int_0^{2\pi} < -y, x > \Big|_{x=\cos t,\ y=\sin t} \cdot < \cos t,\ \sin t >'\ dt \\
&= \int_0^{2\pi} < -\sin t,\ \cos t > \cdot < -\sin t,\ \cos t >\ dt \\
&= \int_0^{2\pi} (\sin^2 t + \cos^2 t)\, dt \\
&= \int_0^{2\pi} 1\, dt \\
&= 2\pi\,.
\end{aligned}$$

We have walked against wind all the way in this round trip!

The same logic applies to any location-dependent multiple of F as long as the symmetry is preserved. For example, the familiar one below qualifies:

$$G(X) = \frac{F(X)}{||X||^2}\,.$$

Even though, as we know, this vector field passes the *Gradient Test*, it has a positive line integral over a circle:

$$W = \oint_C G \cdot dX = \int_a^b G(X(t)) \cdot G'(t)\, dt > 0\,,$$

because the integrand is positive.

The difference between the two outcomes may be explained by the fact that the constant vector field is *gradient*:

$$< p, q >= \nabla f,\ \text{where } f(x,y) = px + qy\,,$$

while the rotation vector field is *not*:

$$< -y, x >\neq \nabla f,\ \text{for any } z = f(x,y)\,.$$

Is there anything special about line integrals of gradient vector fields over curves that aren't closed? We reach the same conclusion for the discrete case: The line integral depends only on the potential function of F. But the latter is an antiderivative of F! We, therefore, speak of an analog of the original Fundamental Theorem of Calculus II from Volume 2 (there will be FTC I later).

Theorem 6.7.4: Fundamental Theorem of Calculus of Gradient Vector Fields II

If on a subset of $\mathbf{R}^n$, *we have* $F = \nabla f$ *and an oriented curve* C *in* $\mathbf{R}^n$ *starts at point* A *and ends at* B, *then*

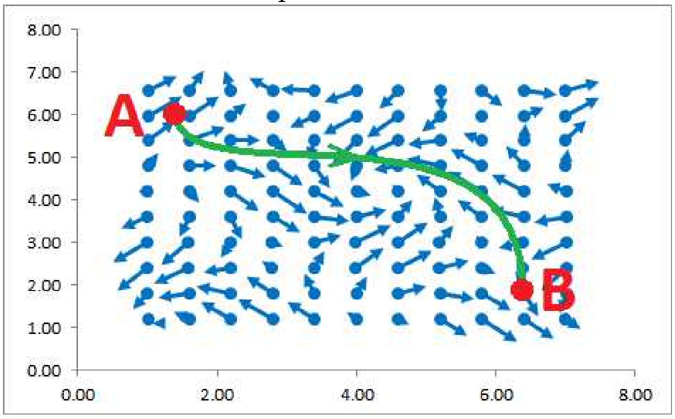
$$\boxed{\int_C F \cdot dX = f(B) - f(A)}$$

Proof.

Suppose we have:

$$F = \nabla f\,,$$

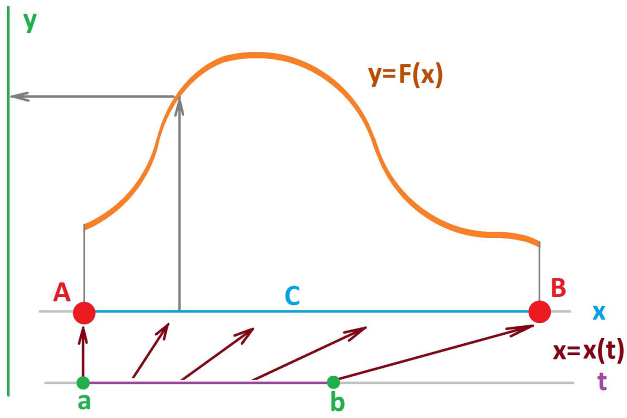
and an oriented curve C in $\mathbf{R}^n$ that starts at point A and ends at B:



Then, after parametrizing C with $X = X(t)$, $a \le t \le b$, we have via the *Fundamental Theorem of Calculus* (Chapter 3IC-1) and the *Chain Rule* (Chapter 2DC-4):

$$\begin{aligned}
W &= \int_C F \cdot dX \\
&= \int_a^b F(X(t)) \cdot X'(t)\, dt \\
&= \int_a^b \nabla f(X(t)) \cdot X'(t)\, dt && \text{We recognize the integrand as a part of CR.} \\
&= \int_a^b \frac{d}{dt} f(X(t))\, dt && \text{We apply now FTC II.} \\
&= f(X(t))\Big|_a^b \\
&= f(X(b)) - f(X(a)) \\
&= f(B) - f(A)\,.
\end{aligned}$$

For dimension $n = 1$, we just take $y = F(x)$ to be a numerical function with antiderivative f and C is the interval $[A, B]$ in the x-axis.



We also choose $x = x(t)$ to be a parametrization of this interval so that $x(a) = A$ and $x(b) = B$. Then we have from above:

$$\int_C F\,dX = \int_{x=A}^{x=B} F\,dx = \int_{t=a}^{t=b} F(x(t))x'(t)\,dt = f(x(t))\Big|_{t=a}^{t=b} = f(x(b)) - f(x(a)) = f(B) - f(A)\,.$$

We have another interpretation of *substitution in definite integrals*.

Not only the proof but also the formula itself looks like the familiar Fundamental Theorem of Calculus for numerical integrals from Volume 3 (Chapter 3IC-1). Because it is restricted to gradient vector fields, this is just a *preliminary version*.

Warning!

Before applying the formula, confirm that the vector field is *gradient*! The example of $F =< -y, x >$ is to be remembered at all times.

So, if F is a gradient vector field then

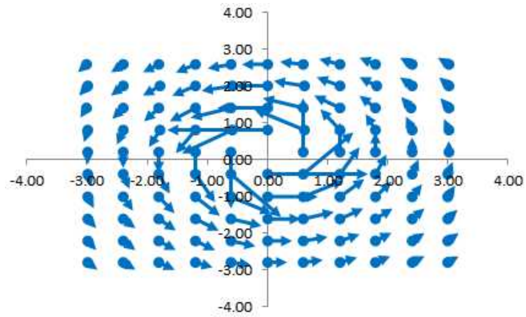
$$\oint_C F \cdot dX = 0\,.$$

Therefore, the work is zero on net so that there is no gain or loss of energy. This is the reason why gradient vector fields are also called *conservative*.

Example 6.7.5: gradient?

Consider this *rotation vector field*, $V =< -y, x >$, and especially its multiple:

$$F = \frac{V}{||V||^2} = \frac{1}{x^2+y^2} < y,\ -x >= \left\langle \frac{y}{x^2+y^2},\ -\frac{x}{x^2+y^2} \right\rangle =< p, q > \ .$$



We previously demonstrated the following:

$$\begin{array}{l} p_y = \dfrac{\partial}{\partial y}\dfrac{y}{x^2+y^2} = \dfrac{1\cdot(x^2+y^2) - y\cdot 2y}{(x^2+y^2)^2} = \dfrac{x^2-y^2}{(x^2+y^2)^2} \\ q_x = \dfrac{\partial}{\partial x}\dfrac{-x}{x^2+y^2} = -\dfrac{1\cdot(x^2+y^2) - x\cdot 2x}{(x^2+y^2)^2} = -\dfrac{y^2-x^2}{(x^2+y^2)^2} \end{array} \implies \operatorname{rot} F = q_x - p_y = 0\,.$$

So, the rotor of the vector field is zero and it passes the *Gradient Test*; however, is it gradient? We demonstrate now that it is not. Indeed, suppose $X = X(t)$ is a counterclockwise parametrization of the circle. Then $F(X(t))$ is parallel to $X'(t)$. Therefore, $F(X(t)) \cdot X'(t) > 0$. It follows that the line integral along the circle is positive:

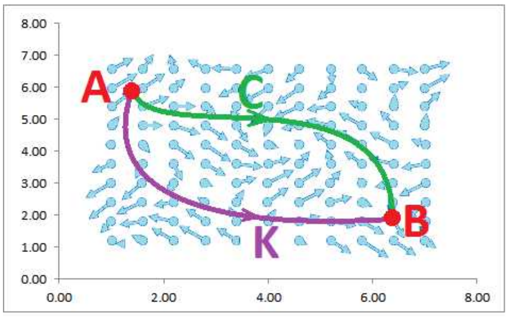
$$0 = \oint_C F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt > 0\,.$$

It is as if we have climbed a spiral staircase! A contradiction.

Not only the expression on right

$$\int_C F \cdot dX = f(B) - f(A)$$

is independent of parametrization of the curve C, it is independent of our choice of C as long as it is from A to B!



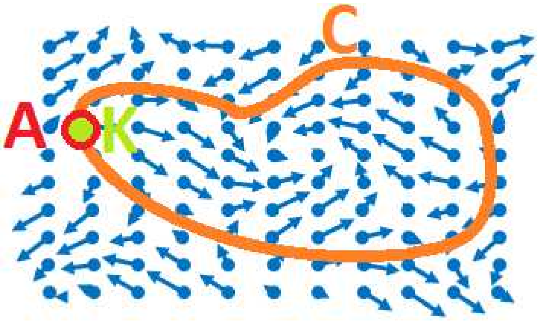
Definition 6.7.6: path-independent vector field

A vector field defined on a subset D of $\mathbf{R}^n$ is called *path-independent* if any line integral along a curve depends only on the start- and the end-points of the curve; i.e.,

$$\boxed{\int_C F \cdot dX = \int_K F \cdot dX}$$

for any two curves C and K from point A to point B that lie entirely in D.

What if $A = B$? What can we say about line integral along a *closed* curve C? As an example, consider this: if we stay home, we don't do any work! We are talking about a constant curve, $K = \{A\}$. Let's compare it to another curve C.



The parametrization of K is trivial: $X(t) = A$ on the whole interval $[a, b]$. Therefore, $X'(t) = 0$ and we have:

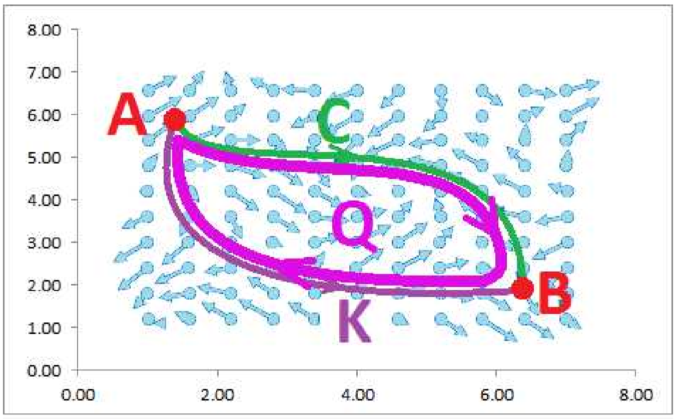
$$\int_C F \cdot dX = \int_K F \cdot dX = \int_a^b F(X(t)) \cdot X'(t)\, dt = \int_a^b F(X(t)) \cdot 0\, dt = 0\,.$$

The converse is also true. Suppose we have two curves C and K from A to B. Just as in the last section,

we create a new, *closed* curve from them. We glue C and the reversed K together:

$$Q = C \cup -K\,.$$

It goes from A to A.



Then, from *Additivity* and *Negativity* we have:

$$0 = \int_Q F \cdot dX = \int_C F \cdot dX + \int_{-K} F \cdot dX = \int_C F \cdot dX - \int_K F \cdot dX\,.$$

Therefore,

$$\int_C F \cdot dX = \int_K F \cdot dX\,.$$

In summary, we have the following.

Theorem 6.7.7: Path-independent Over Loops

A vector field defined on a subset D of $\mathbf{R}^n$ is path-independent if and only if it has all of its line integrals along closed curves in D equal to zero.

We have established the following.

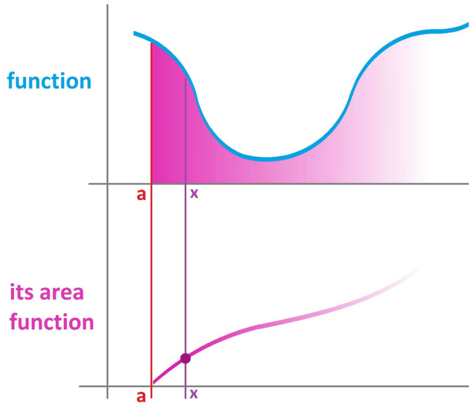
Theorem 6.7.8: Gradient => Path-independent

All gradient vector fields are path-independent.

Proof.

- F is a gradient vector field, then
- $\oint_C F \cdot dX = 0$, then
- F is path-independent.

Recall that we considered Riemann integral (the area under the graph) but with a *variable upper limit*. It is illustrated below: x runs from a to b and beyond.



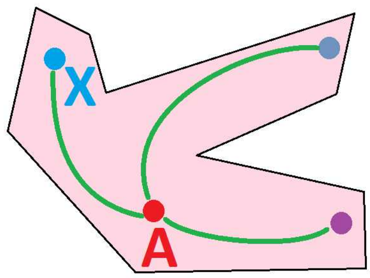
Then the *Fundamental Theorem of Calculus II* states that for any continuous function F on $[a,b]$, the function defined by

$$\int_a^x F\,dx$$

is an antiderivative of F on (a,b). In the new setting, we have a path-independent vector field F defined on some set D in $\mathbf{R}$ and we need to find its potential function, i.e., a function the gradient of which is F, $\nabla f = F$. First, we choose an arbitrary point A in D and then do a lot of line integration. We define for each X in D:

$$f(X) = \int_C F \cdot dX\,,$$

where C is any curve from A to X. A choice of C doesn't matter because F is path-independent by assumption.



There is an extra requirement.

Theorem 6.7.9: Fundamental Theorem of Calculus of Gradient Vector Fields I

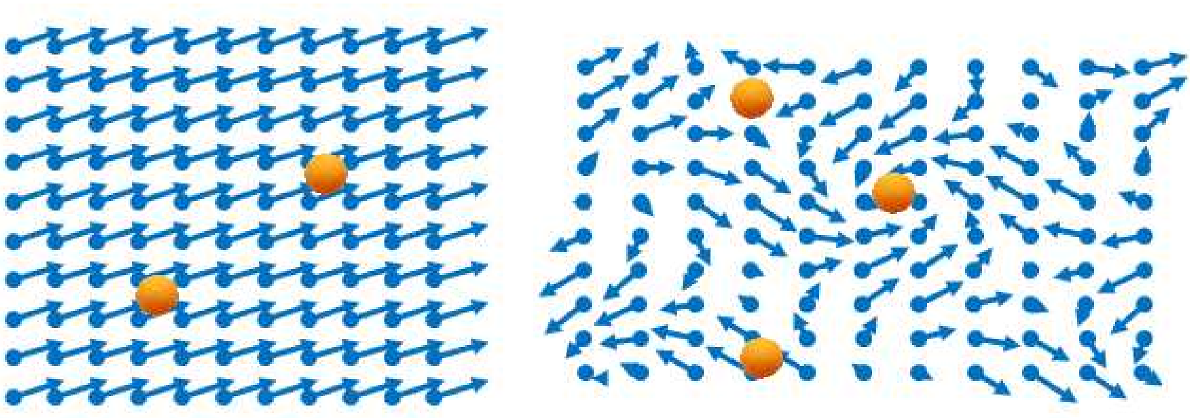
For any gradient vector field F defined on a path-connected region in $\mathbf{R}^n$, the function defined for a fixed A in D by:

$$\boxed{f(X) = \int_C F \cdot dX}$$

where C is any curve from A to X within D, is a potential function of F on D.

6.8. How a ball is spun by the stream

Suppose we have a vector field that describes the velocity field of a fluid flow. Let's place a ping-pong ball within the flow. We put it on a pole so that the ball remains fixed while it can freely rotate:



We see the particles bombarding the ball and think of the vector field of the flow as a *force field*. Due to the ball's rough surface, the fluid flowing past it will make it spin around the pole.

It is clear that a constant vector (left) field will produce no spin or rotation. However, it is not the rotation of the vectors that we will speak of. We are not asking: Is a specific particle of water making a circle? Rather:

▶ Does the combined motion of the particles make the ball spin?

For example, this is what we see in the image on right:

- The ball in the center is in the middle of a whirl and will be clearly spun in the counterclockwise direction.
- The ball at the bottom is in the part of the stream with a constant direction but not magnitude. Will it spin?
- The ball at the top is being pushed in various directions at the same time and its spin seems very uncertain.

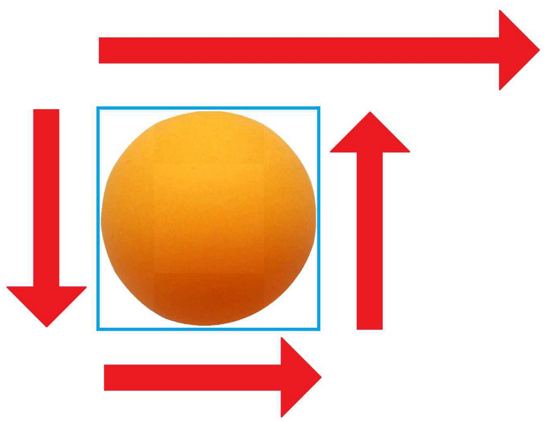
How do we predict and how do we measure the amount of rotation?

Example 6.8.1: ping-pong

The answer is simple when the force is applied to just one side of the ball as in the case of all racket sports:

However, the idea might be utilized for the case of a ball in the stream.

For simplicity, let's assume that we can detect only four distinct values of the vector field on the four sides (on the grid) of the ball.



We also assume at first that these four vectors are tangent to the surface of the ball. In other words, this is just a *vector field*. What is the net effect of these forces on the ball? Think of the ball as a tiny *wind-mill* with the four forces pushing (or pulling) its four blades. We just go around the ball (counterclockwise starting at the bottom) adding these numbers:

$$1 + 1 - 2 + 1 = 1 > 0\,.$$

The ball will spin counterclockwise!

In order to measure the amount of spin, let's assume that this is a *unit* square. Then, of course, the sum above is just a line sum from Chapter 6 representing the *work* performed by the force of the flow to spin the ball.

Let's look at this quantity from the coordinate point of view. We observe that the forces with the same direction but on the opposite sides are cancelled. We see this effect if we re-arrange the terms:

$$W = \text{horizontal: } 1 - 2 \ + \ \text{vertical: } 1 + 1\,.$$

We then represent each vector in terms of its x and y components:

$$\text{force} \ = \ \begin{array}{|ccc|}\hline \bullet- & \to\to & -\bullet \\ | & & | \\ \downarrow & & \uparrow \\ | & & | \\ \bullet- & \to & -\bullet \\ \hline \end{array} \ = \ \begin{array}{|ccc|}\hline \bullet- & 2 & -\bullet \\ | & & | \\ -1 & & 1 \\ | & & | \\ \bullet- & 1 & -\bullet \\ \hline \end{array}$$

The expression can then be seen as:

- ▶ $W = -$(the vertical change of the horizontal values) + (the horizontal change of the vertical values).

According to the *Exactness Test* for dimension 2, a function G defined on the edges of a partition is *not* exact when $\Delta_y p \neq \Delta_x q$. We form the following function to study this further.

Definition 6.8.2: difference

For a function G defined on the edges of a partition of the xy-plane, the *difference* of G is a function of two variables defined at the 2-cells of the partition and denoted as follows:

$$\boxed{\Delta G = \Delta_x q - \Delta_y p}$$

where p and q are the x- and y-components of G (i.e., its values on the horizontal and vertical edges respectively).

It is as if we cover the whole stream with those little balls and study their rotation.

Definition 6.8.3: closed function

If its difference is zero, a function defined on the edges of a partition is called *closed*.

The negative rotation simply means rotation in the opposite direction.

Example 6.8.4: rotate without rotation

All vector fields have vectors that change directions, i.e., *rotate*. What if they don't? Let's consider a

flow with a constant direction but variable magnitude:

$$
\text{Force} \ = \ \begin{array}{|ccc|}
\hline
\bullet- & \to\to & -\bullet \\
| & & | \\
\cdot & & \cdot \\
| & & | \\
\bullet- & \to & -\bullet \\
\hline
\end{array}
\ = \ \begin{array}{|ccc|}
\hline
\bullet- & 2 & -\bullet \\
| & & | \\
0 & & 0 \\
| & & | \\
\bullet- & 1 & -\bullet \\
\hline
\end{array}
$$

The rotor is -1, but where is rotation? Well, the speed of the water on one side is faster than on the other and this difference is the cause of the ball's spinning.

With this new concept, we can restate the Exactness Test.

Theorem 6.8.5: Exactness Test Dimension 2

If G is exact, it is closed; briefly:

$$\boxed{\Delta(\Delta h) = 0}$$

for all h.

Let's try a more general point of view: vector fields.

Example 6.8.6: computation

We represent each vector in terms of its x- and y-components:

$$
\text{Force} \ = \ \begin{array}{|ccc|}
\hline
\bullet- & \to\to & -\bullet \\
| & & | \\
\downarrow & & \uparrow \\
| & & | \\
\bullet- & \to & -\bullet \\
\hline
\end{array}
\ = \ \begin{array}{|ccc|}
\hline
\bullet- & < 2, 0 > & -\bullet \\
| & & | \\
< 0, -1 > & & < 0, 1 > \\
| & & | \\
\bullet- & < 1, 0 > & -\bullet \\
\hline
\end{array}
$$

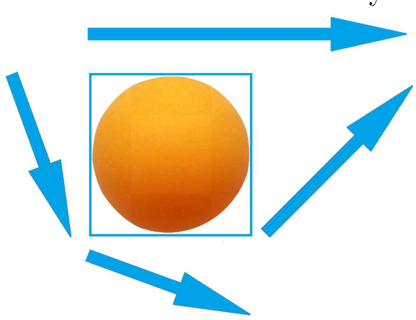
The expression can then be seen as:

- ► $W = -$(the vertical change of the horizontal vectors) $+$ (the horizontal change of the vertical vectors);

or:

$$W = \text{horizontal: } \ -(2-1) + \ \ \text{vertical: } 1 - (-1)\,.$$

Of course, only the vertical/horizontal components of the vectors acting along the vertical/horizontal edges matter! So the result should remain the same if we modify make the other components non-zero:



Then, we have:

$$F = \boxed{\begin{array}{lcr} \bullet- & <2,0> & -\bullet \\ | & & | \\ <1/2,-1> & & <1,1> \\ | & & | \\ \bullet- & <1,-1> & -\bullet \end{array}}$$

The value of W above remains the same even though the forces are directed off the tangent of the ball! The difference is between a real-valued 1-form and a vector-valued 1-form.

If $F =< p, q >$, we have the following componentwise:

$$p = \boxed{\begin{array}{lcr} \bullet- & 2 & -\bullet \\ | & & | \\ 1/2 & & 1 \\ | & & | \\ \bullet- & 1 & -\bullet \end{array}} \implies \Delta_y p = 2 - 1 = 1, \quad q = \boxed{\begin{array}{lcr} \bullet- & 0 & -\bullet \\ | & & | \\ -1 & & 1 \\ | & & | \\ \bullet- & -1 & -\bullet \end{array}} \implies \Delta_x y = 1 - (-1) = 2\,.$$

Then,

$$W = -\Delta_y p + \Delta_x q = -1 + 2 = 1\,.$$

This is the familiar *rotor* from before!

Here is another way to arrive to this quantity. If C is the border of the square oriented in the counterclockwise direction, the line sum along C gives us the following:

$$\begin{aligned} W &= \sum_C F \\ &= \boxed{\begin{array}{ccc} & <2,0>\cdot<-1,0> & + \\ <1/2,-1>\cdot<0,-1> & + & <1,1>\cdot<0,1> \\ + & <1,-1>\cdot<1,0> & \end{array}} \\ &= \boxed{\begin{array}{ccc} & -2 & + \\ 1 & + & 1 \\ + & 1 & \end{array}} \\ &= 1\,. \end{aligned}$$

According to the *Gradient Test* for dimension 2, a vector field $F =< p, q >$ is *not* gradient when

$$\frac{\Delta p}{\Delta y} \neq \frac{\Delta q}{\Delta x}\,.$$

We form the following function of two variables to study this further.

Definition 6.8.7: rotor

For a vector field F defined on the secondary nodes (the 1-cells) of a partition of region in the xy-plane, the *rotor* of F is a function defined on tertiary nodes (the 2-cells) of the partition and denoted as follows:

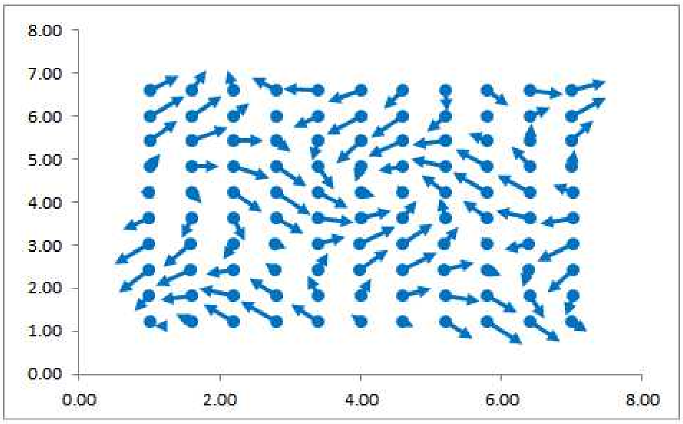
$$\boxed{\operatorname{rot} F = \frac{\Delta p}{\Delta y} - \frac{\Delta q}{\Delta x}}$$

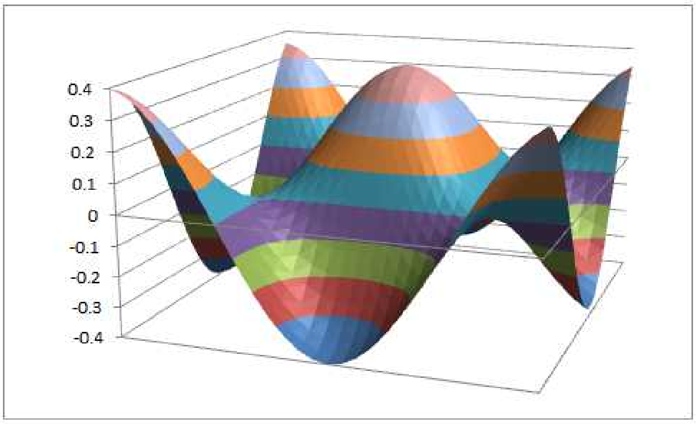
where p and q are the x- and y-components of V (i.e., its values on the horizontal and vertical edges respectively).

Definition 6.8.8: irrotational vector field

If the rotor is zero, the vector field is called *irrotational*.

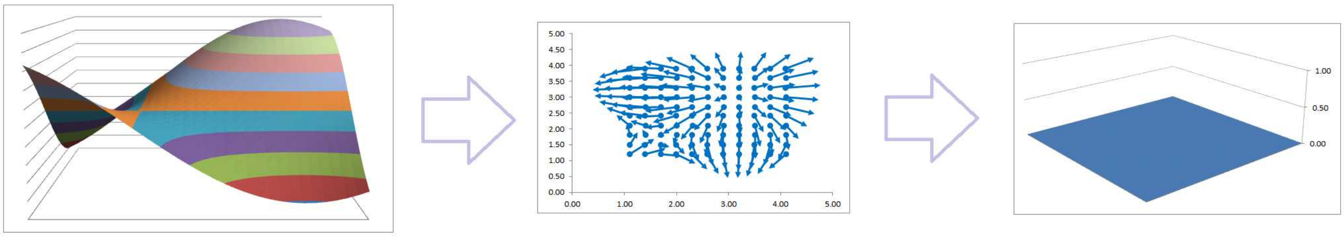
One can see a high value of the rotor in the center and zero around it in the following example:





Example 6.8.9: rotor of gradient

From the equality of the mixed partial difference quotients, it follows that the rotor of the gradient of a function gives values exactly equal to 0:



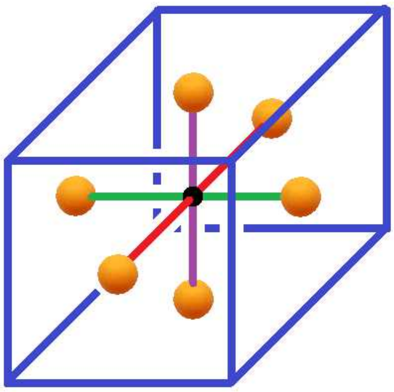
With this new concept, we can restate the Gradient Test.

Corollary 6.8.10: Gradient Test For Dimension 2

If a vector field is gradient, then it's irrotational.

What about the 3-*dimensional* space?

Once again, we place a small ball within the flow in such a way that the ball remains fixed while being able to rotate. If the ball has a rough surface, the fluid flowing past it will make it spin. In the discrete case, each face, i.e., a 2-cell, of the partition is subject to the 2-dimensional analysis presented above. In other words, the ball located within a face rotates around the axis perpendicular to the face:



According to the *Exactness Test* for dimension 3, $G =< p, q, r >$ is *not* exact when one of these fails:

$$\Delta_y p = \Delta_x q,\ \Delta_z q = \Delta_y r,\ \Delta_x r = \Delta_z p\,.$$

We form the following vector field to study this further.

Definition 6.8.11: difference

For a function F defined on the secondary nodes (edges) of a partition of the xyz-space, the *difference* of G is a function defined at the tertiary nodes (2-cells) of a partition of a cell in the xyz-space and denoted as follows:

$$\boxed{\Delta G = \begin{cases} \Delta_y r - \Delta_z q & \text{on the faces parallel to the } yz\text{-plane} \\ \Delta_z p - \Delta_x r & \text{on the faces parallel to the } xz\text{-plane} \\ \Delta_x q - \Delta_y p & \text{on the faces parallel to the } xy\text{-plane} \end{cases}}$$

where p, q, and r are the x-, y-, and z-components of G respectively. If the difference is zero, G is called *closed*.

Of course, the 3-dimensional difference is made of the three 2-dimensional ones with respect to each of the three pairs of coordinates.

Theorem 6.8.12: Exactness Test For Dimension 3

If G is exact, it is closed; briefly:

$$\boxed{\Delta(\Delta h) = 0}$$

for all h.

Same statement as for dimension 2!

According to the *Gradient Test* for dimension 3, a vector field $V =< p, q, r >$ is *not* gradient when one of these fails:

$$\frac{\Delta p}{\Delta y} = \frac{\Delta q}{\Delta x},\ \frac{\Delta q}{\Delta z} = \frac{\Delta r}{\Delta y},\ \frac{\Delta r}{\Delta x} = \frac{\Delta p}{\Delta z}\,.$$

Definition 6.8.13: curl

For a function F defined on the edges of a partition of the xyz-space, the *curl*, of F is a function of three variables defined at the 2-cells of a partition of a cell

in the xyz-space and denoted as follows:

$$\operatorname{curl} F = \begin{cases} \dfrac{\Delta r}{\Delta y} - \dfrac{\Delta q}{\Delta z} & \text{on the faces parallel to the } yz\text{-plane} \\ \dfrac{\Delta p}{\Delta z} - \dfrac{\Delta r}{\Delta x} & \text{on the faces parallel to the } xz\text{-plane} \\ \dfrac{\Delta q}{\Delta x} - \dfrac{\Delta p}{\Delta y} & \text{on the faces parallel to the } xy\text{-plane} \end{cases}$$

where p, q, and r are the x-, y-, and z-components of F respectively. If the curl is zero, F is called *irrotational*.

Of course, the curl is made of the three rotors with respect to the three pairs of coordinates.

Corollary 6.8.14: Gradient is Irrotational

If a vector field is gradient, then it's irrotational.

Same statement!

The two theorems can be restated in an even more concise form, in terms of the compositions of these *functions of functions*:

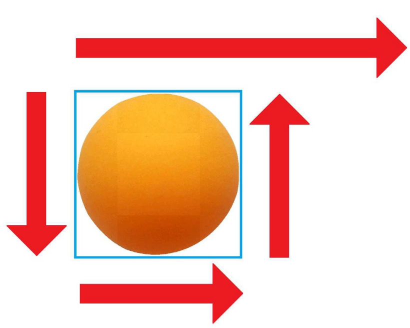
$$\boxed{\Delta\Delta = 0}$$

When no secondary nodes are specified, we deal with discrete forms. Then, if travel along the following diagram, we end up at *zero* no matter what the starting point is:

$$\text{0-forms} \xrightarrow{\ \Delta\ } \text{1-forms} \xrightarrow{\ \Delta\ } \text{2-forms.}$$

6.9. The Fundamental Theorem of Discrete Calculus of degree 2

Suppose curve C is the border of the rectangle R oriented in the counterclockwise direction.



Suppose the flow is given by these numbers as defined on each of the edges of the rectangle:

$$G = \begin{array}{|ccc|} \hline \bullet & p_3 & \bullet \\ q_4 & & q_2 \\ \bullet & p_1 & \bullet \\ \hline \end{array} \quad C = \begin{array}{|ccc|} \hline \bullet & \leftarrow & \bullet \\ \downarrow & & \uparrow \\ \bullet & \to & \bullet \\ \hline \end{array}$$

Then the line integral along C is the following:

$$\begin{array}{ll} W &= \sum_C G \\ &= \begin{array}{ccc} & & -p_3 \\ + & -q_4 & + \quad q_2 \\ + & & p_1 \end{array} \\ &= -p_3 - q_4 + q_2 + p_1 \\ &= (q_2 - q_4) - (p_3 - p_1) \quad \text{This is the horizontal change of } q \text{ and vertical change of } p. \\ &= \Delta_x G - \Delta_y G \\ &= \Delta G\,. \end{array}$$

As you can see, rearranging the four terms of the work that come from the trip around the square creates the following. First, it is the difference of the vertical flow on the two sides of the ball and, second, it is the difference of the horizontal flow on the other two sides. Finally, the difference of these two quantities appears and it indicates the total flow. It is the *difference* of G.

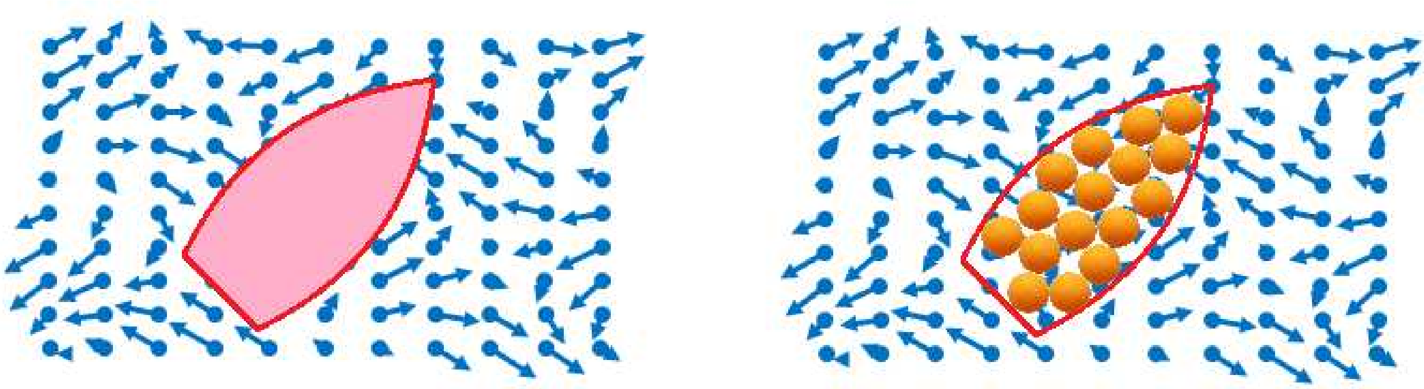
We have a preliminary result below.

Theorem 6.9.1: Sum Around Cell

In a partition of a plane region R, if C is a simple closed curve that constitutes the boundary of a single 2-cell D of the partition by going counterclockwise around D, we have the following for any function G defined on the secondary nodes of the partition:

$$\boxed{\sum_C G = \Delta G}$$

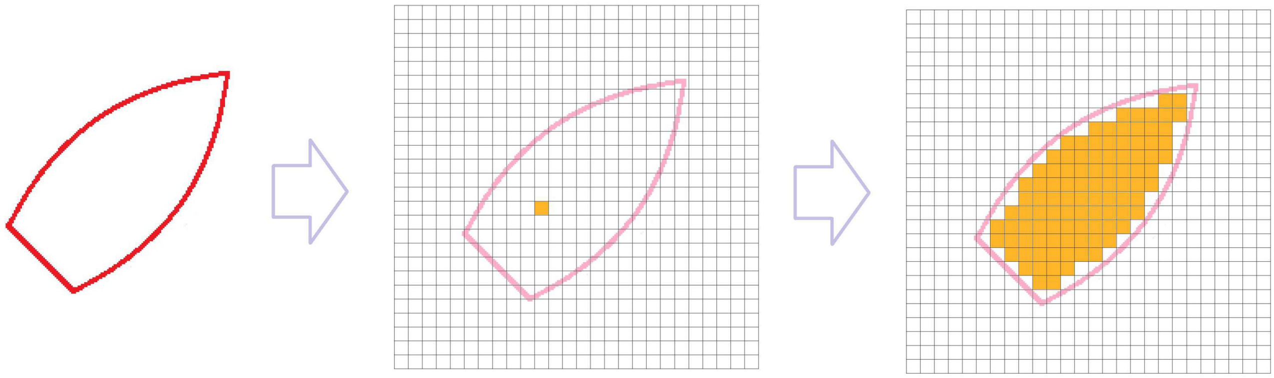
What if we have a more complex object in the stream? For example, a boat:



How do we measure the amount of flow around it?

We approach the problem as follows: We imagine that there are many little balls in the flow forming the shape we are considering and then find the amount of the flow around the balls. Note that every ball will try to rotate all of its adjacent balls in the same direction at the same speed with no more flow required. This idea of cancellation of spin takes an algebraic form below.

We will start with a single rectangle and then build more and more complex regions on the plane from the rectangles of our grid – as if each contains a ball – while maintaining the formula:

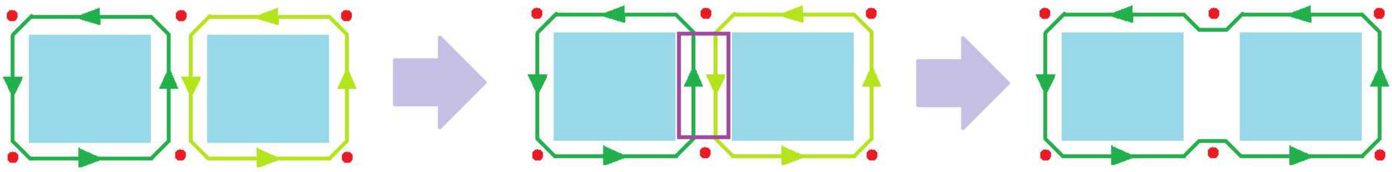


Let's put two rectangles together. Suppose we have two adjacent ones, R_1 and R_2, bounded by curves C_1 and C_2. We write the Fundamental Theorem for either and then add the two:

$$\begin{array}{rrcl} & \sum_{C_1} G & = & \sum_{R_1} \Delta G \\ + & & & \\ & \sum_{C_2} G & = & \sum_{R_2} \Delta G \\ \hline & \sum_{C_1 \cup C_2} G & = & \sum_{R_1 \cup R_2} \Delta G \end{array}$$

In the right-hand side, we have a single sum according to *Additivity* of sums and in the left-hand side, we have a single sums according to *Additivity*. Here $C_1 \cup C_2$ is the curve that consists of C_1 and C_2 traveled consecutively.

Now, this is an unsatisfactory result because $C_1 \cup C_2$ doesn't bound $R_1 \cup R_2$. Fortunately, the left-hand side can be simplified: the two curves share an edge but travel it in the *opposite* directions.

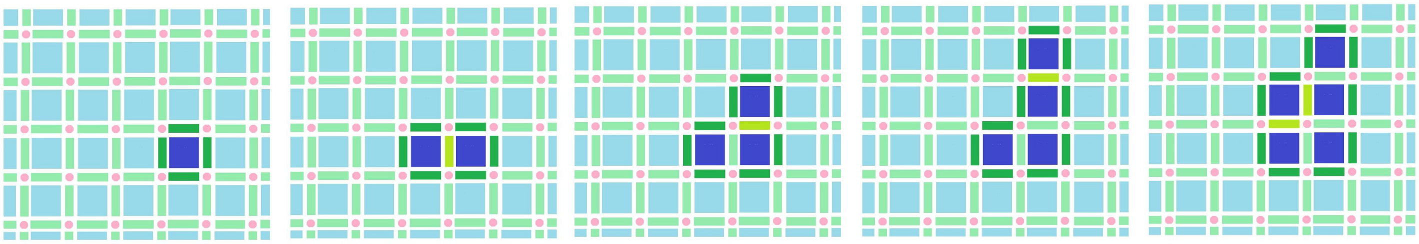


We have a cancellation according to *Negativity* for sums. The result is:

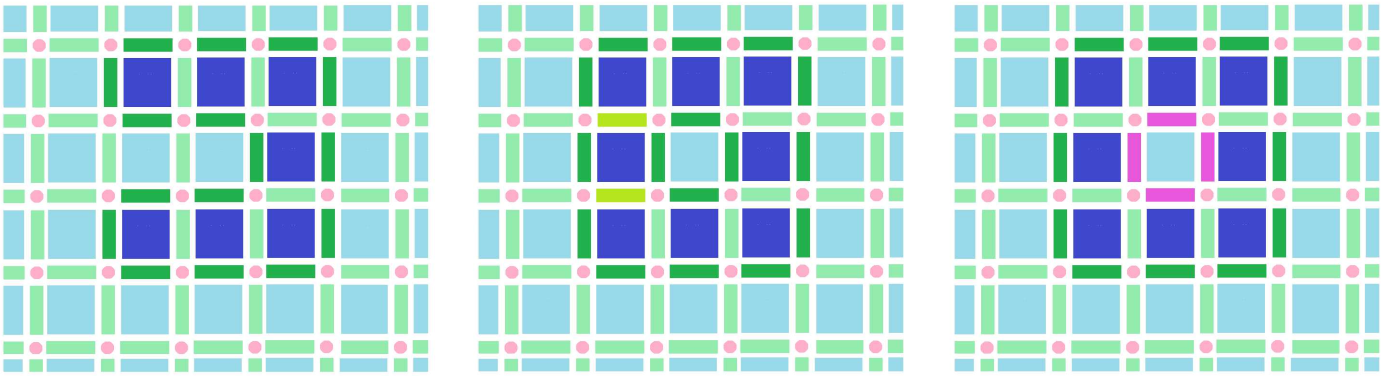
$$\sum_{\partial D} \Delta G = \sum_{D} \Delta G\,,$$

where D is the union of the two rectangles and ∂D is its boundary. We have constructed the Fundamental Theorem for this more complex region!

We continue on adding one rectangle at a time to our region D and cancelling the edges shared with others producing bigger and bigger curve $C = \partial D$ that bounds D:



We can add as many rectangles as we like and producing larger and larger region made of the rectangles and bounded by a single closed curve made of edges... unless we circle back!



Then the boundary curve might break into two...

We will ignore this possibility for now and state the second preliminary version of the main theorem.

Theorem 6.9.2: Sum Around Region

In a partition of a plane region R, if C is a simple closed curve that constitutes the boundary of R by going counterclockwise around R, we have for any function G defined on the secondary nodes of the partition:

$$\boxed{\sum_{C=\partial R} G = \sum_R \Delta G}$$

What if the function is the difference? Then its difference is zero and, therefore, our formula takes this form:

$$\sum_{\partial D} \Delta G = \sum_D \Delta G = \sum_D 0 = 0\,.$$

The sum along any closed curve is then zero and, according to the *Path-independence Theorem*, G is path-independent. Then,

$$\sum_C \Delta G = f(B) - f(A)\,,$$

for any curve C from A to B, where f is a potential function of G. We have arrived at the Fundamental Theorem of Calculus for differences. It follows that the Fundamental Theorem is its generalization. However, as the Fundamental Theorem of Calculus for parametric curves, i.e., degree 1, indicates, *there are more than one fundamental theorem for each dimension*!

What is our function doesn't depend on y, i.e., $G(x,y) = q(x)$, while R is a rectangle $[a,b]\times[c,d]$? In the left-hand side of the formula, the sums along the two horizontal sides of R cancel each other:

$$\sum_C G \cdot dX = q(b) - q(b)\,.$$

In the right-hand side of the formula, we have:

$$\sum_R \Delta G = \sum_{[a,b]\times[c,d]} \Delta_x G = \sum_{[a,b]} \Delta q\,.$$

We have arrived at the original *Fundamental Theorem of Discrete Calculus* (degree 1) from Volume 2 (Chapter 3IC-1):

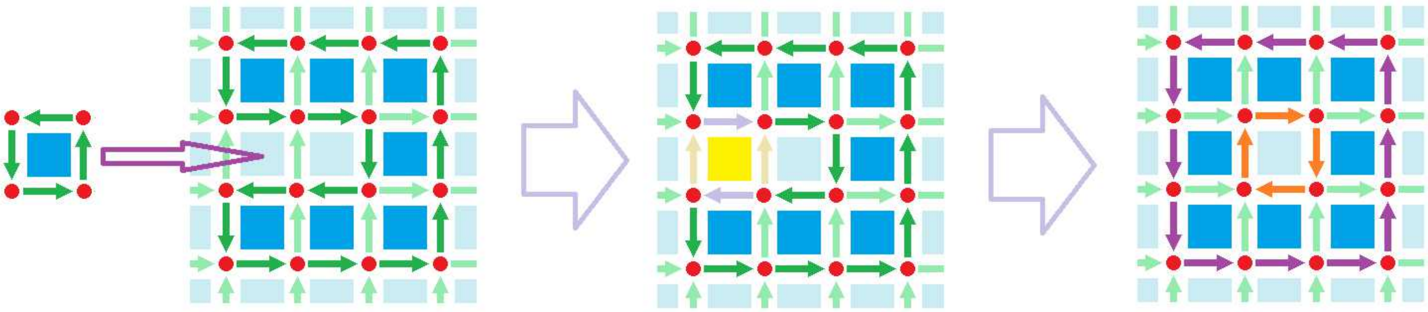
$$q(b) - q(a) = \sum_{[a,b]} \Delta q\,.$$

Not only have we derived the degree 1 from degree 2, but also both theorems have the same form! We realize that in the above formula,

$$\sum_{\{a,b\}} q = \sum_{[a,b]} \Delta q\,,$$

the right-hand side is an integral of a 1-form over a (1-dimensional) region, $R = [a, b]$, while the left hand-side is a 0-form over the boundary, $\partial R = \{a, b\}$, properly oriented, of that region.

Now, what if the boundary curve does break into two when we add a new square? In the example below the square is added along with four of its edges. As a result, we add the two vertical edges while the two horizontal are cancel as before. Thus a new square is seamlessly added but we also see the appearance of a *hole*:

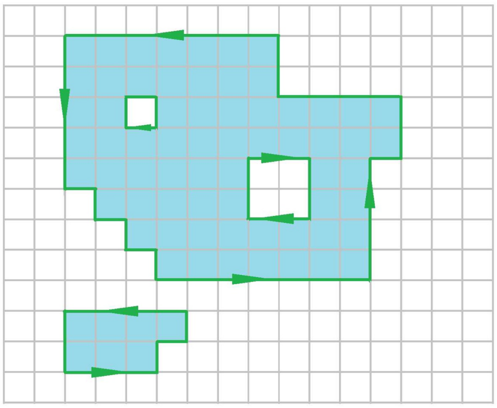


The difference is dramatic: not only the boundary of the region is now made of two curves but also the one outside goes counterclockwise (as before) while the one inside goes clockwise! However, either curve has the region to its *left*.

Our formula,

$$\sum_{C} G = \sum_{R} \Delta G\,,$$

doesn't work anymore, even though the meaning of the right-hand side is still clear. But what should be the meaning of the left-hand side? It should be the total sum of G over all boundary curves of R, correctly oriented!



Thus, the fundamental is the relation between a region R in a partition and its boundary ∂R.

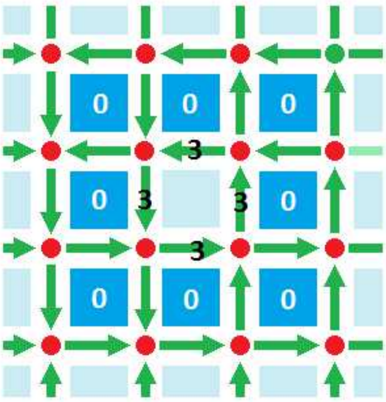
Theorem 6.9.3: Fundamental Theorem of Discrete Calculus of Degree 2

In a partition of a plane region R, we have the following for any function G defined on the secondary nodes of the partition:

$$\boxed{\sum_{\partial R} G = \sum_{R} \Delta G}$$

Example 6.9.4: loop is boundary

We know that for a region bounded by a simple closed curve, the sum along *any* closed curve is 0. Let's take a look at what happens in regions with *holes*. Consider this *rotation* function G:



Its values are ± 1 with directions indicated except for the four edges in the middle with values of ± 3. The function is defined on the 3×3 region R that excludes the middle square. By direct examination we show that the difference of G is zero at every face of R:

$$\Delta G = 0\,.$$

So, G passes the *Exactness Test*; however, is it exact? We demonstrate now that it is not. Indeed, the sum of G along the outer boundary of R isn't zero:

$$\sum_C G = 12\,.$$

How does it work with our theorem:

$$\sum_C G = \sum_R \Delta G\,?$$

It seems that the left-hand side is positive while right-hand side is zero... What we have overlooked is that G and, therefore, its difference are undefined at the middle square! So, C doesn't bound R. In fact, the boundary of R includes another curve, C', going clockwise. Then,

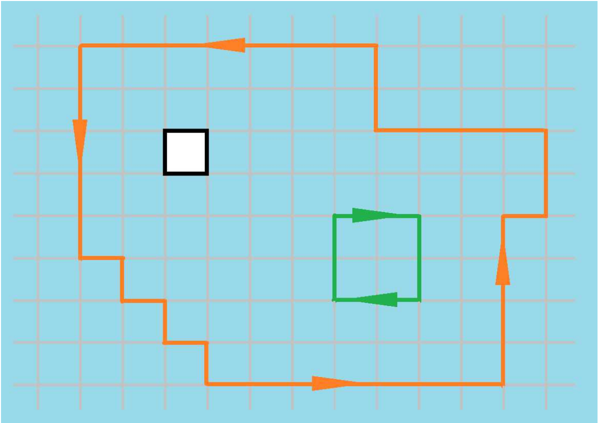
$$\sum_C G + \sum_{C'} G = \sum_R \Delta G = 0\,.$$

Therefore, we have:

$$\sum_C G = \sum_{-C'} G\,.$$

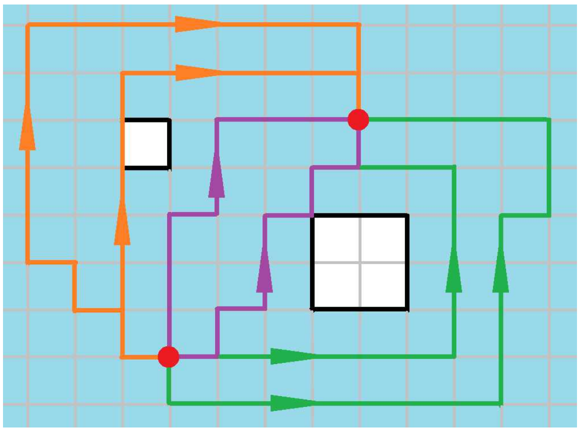
So, moving from the larger path to the smaller (or vice versa) doesn't change the sum! Also notice that the sums from one corner to the opposite are 6 and -6. There is no path-independence!

To summarize, even when the difference is – within the region – zero, the sum along a path that *goes around the hole* may be non-zero:



Furthermore, the sum remains the same for all closed curves as long as they make exactly the same number

of turns around the origin! The meaning of path-independence changes accordingly; it all depends on how the curve goes between the holes:



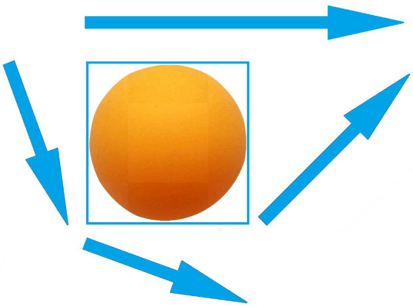
Next, we consider the relation of this line integral that represents the *work* performed by the flow to spin the ball and the rotor of the vector field.

Recall that we have a vector field $F =< p, q >$ of the velocity field of a fluid flow with a ping-pong ball within it that can freely rotate but not move. We measure the amount of rotation as the work performed by the force of the flow rotating the ball.

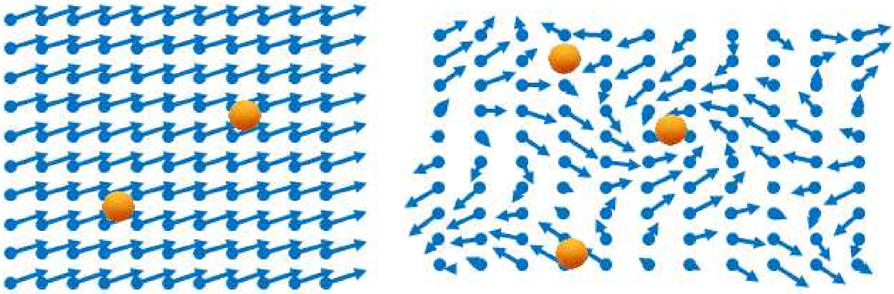
Let's first suppose we have a grid on the plane with rectangles:

$$\Delta x \times \Delta y\,.$$

Suppose that the flow rotates this rectangle just like the ball before.



We can also look at a vector field $F =< p, q >$ of the velocity field of a fluid flow.



Thus, the Riemann sum of the vector field along the boundary of a rectangle is equal to the (double) Riemann sum of the rotor over this rectangle and, furthermore, over any region made of such rectangles.

Theorem 6.9.5: Fundamental Theorem of Discrete Calculus for Vector Fields

In a partition of a plane region R, we have the following, we have for any vector

field F defined on the secondary nodes of the partition:

$$\boxed{\sum_{\partial R} F \cdot \Delta X = \sum_{R} \operatorname{rot} F \, \Delta A}$$

Proof.

The proof can be independent from the last theorem. Suppose curve C is the border of the rectangle R oriented in the counterclockwise direction. Suppose the vector field is given by these vectors as defined on each of the edges of the rectangle, which is shown on right:

$$F = \boxed{\begin{array}{ccc} \bullet & <p_3, q_3> & \bullet \\ <p_4, q_4> & & <p_2, q_2> \\ \bullet & <p_1, q_1> & \bullet \end{array}} \quad \Delta X = \boxed{\begin{array}{ccc} \bullet & <-\Delta x, 0> & \bullet \\ <0, -\Delta y> & & <0, \Delta y> \\ \bullet & <\Delta x, 0> & \bullet \end{array}}$$

Then the Riemann sum along C is:

$$\begin{aligned} W &= \sum_C F \cdot \Delta X \\ &= \begin{array}{ccc} & <p_3, q_3> \cdot <-\Delta x, 0> & \\ + <p_4, q_4> \cdot <0, -\Delta y> & + & <p_2, q_2> \cdot <0, \Delta y> \\ + & <p_1, q_1> \cdot <\Delta x, 0> & \end{array} \\ &= \begin{array}{ccc} & -p_3 \Delta x & \\ + \quad -q_4 \Delta y & + & q_2 \Delta y \\ + & p_1 \Delta x & \end{array} \\ &= -p_3 \Delta x - q_4 \Delta y + q_2 \Delta y + p_1 \Delta x \\ &= (q_2 - q_4) \Delta y - (p_3 - p_1) \Delta x \\ &= \frac{q_2 - q_4}{\Delta x} \Delta x \Delta y - \frac{p_3 - p_1}{\Delta y} \Delta y \Delta x \\ &= \left(\frac{\Delta q}{\Delta x} - \frac{\Delta p}{\Delta y} \right) \Delta x \Delta y \, . \end{aligned}$$

6.10. Green's Theorem: the Fundamental Theorem of Calculus for vector fields in dimension 2

According to the *Gradient Test* for dimension 2, a vector field $F = <p, q>$ is *not* gradient when $p_y \neq q_x$. We form the following function of two variables to study this further (as if we cover the whole stream with those little balls).

Definition 6.10.1: rotor

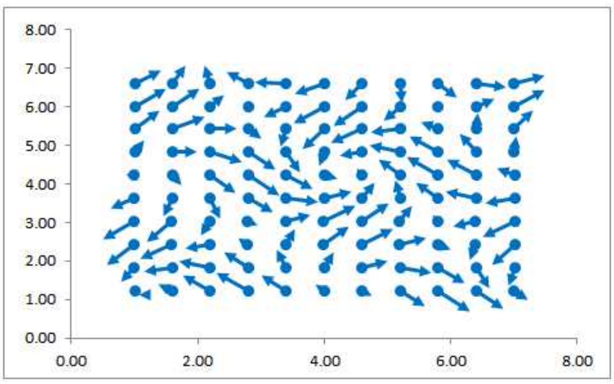
The *rotor* of a differentiable on an open region on the plane vector field $F =< p, q >$ is a function of two variables defined on the region and denoted as follows

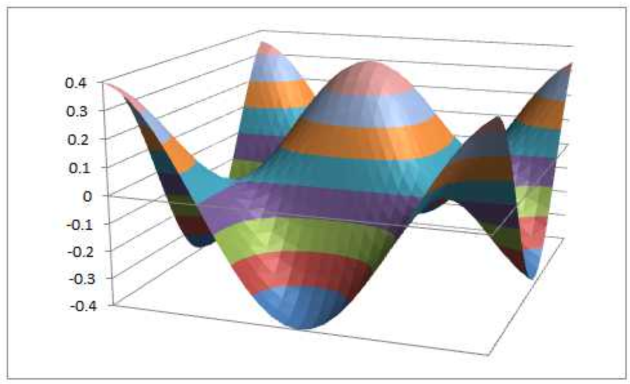
$$\boxed{\operatorname{rot} F = q_x - p_y}$$

Definition 6.10.2: irrotational vector field

If the rotor is zero, the vector field is called *irrotational*.

One can see a high value of the rotor in the center and zero around it in the following example:



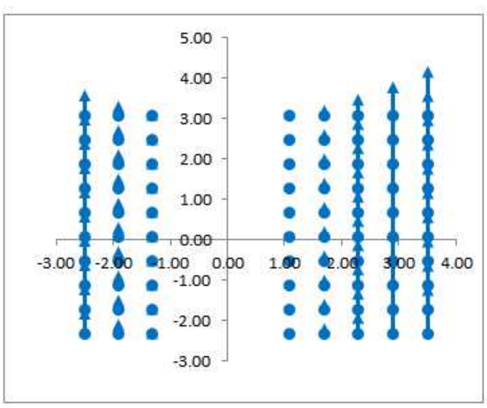


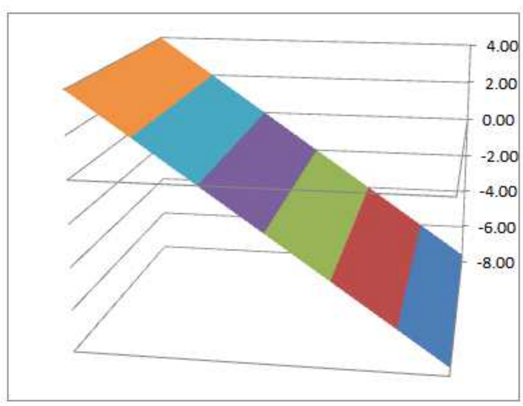
The negative rotation simply means rotation in the opposite direction.

Example 6.10.3: rotating without rotation

All vector fields have vectors that change directions, i.e., "rotate". What if they don't? Let's consider a vector field with a constant direction but variable magnitude. Let's try:

$$F(x, y) =< y^2, 0 > \ .$$

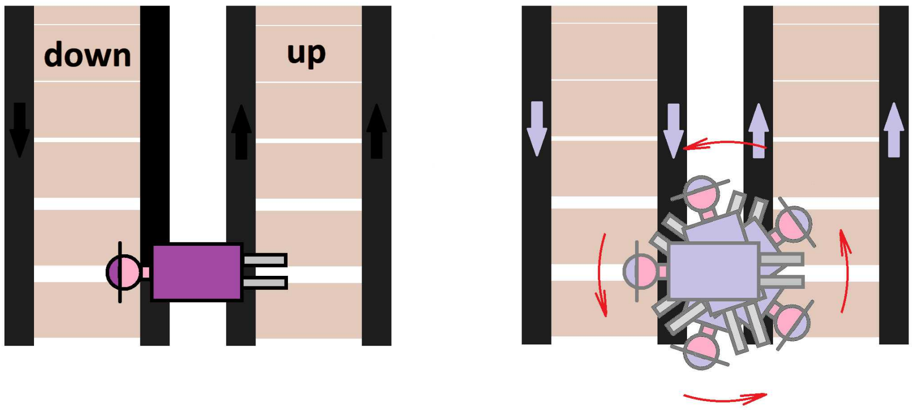




Then

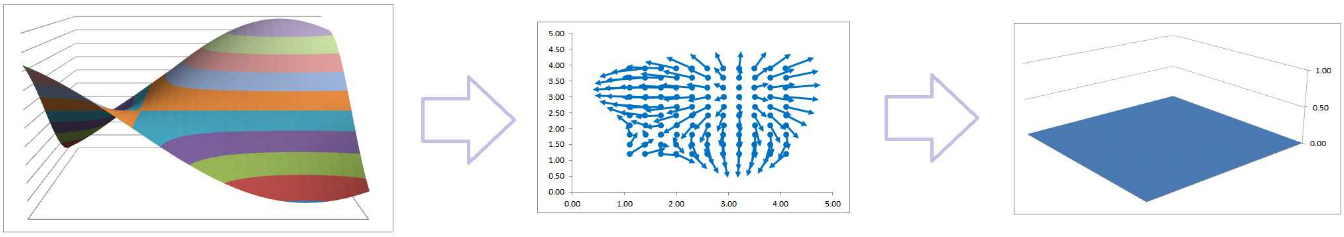
$$\operatorname{rot} V = q_x - p_y = 0 - 2y \neq 0\,.$$

The rotation is again non-zero. In fact the graph of the rotor shows that the rotation will be counterclockwise on right and clockwise on left. The effect is seen when a person lies on the top of two adjacent – up and down – escalators:



Example 6.10.4: rotor of gradient

From the equality of the mixed partial difference quotients, it follows that the rotor of the gradient of a function gives values exactly equal to 0:



With this new concept, we can restate the *Gradient Tests* from Volume 3 (Chapter 3IC-1).

Theorem 6.10.5: Gradient Test For Dimension 2

Suppose F is a vector field on an open region in $\mathbf{R}^2$ with continuously differentiable component functions. If F is gradient (i.e., $F = \operatorname{grad} h$), then it's irrotational: $\operatorname{rot} F = 0$*; briefly:*

$$\boxed{\operatorname{rot}(\operatorname{grad} h) = 0}$$

What about 3-dimensional vector fields? Once again, suppose we have a vector field that describes the velocity field of a fluid flow. We place a small ball within the flow in such a way that the ball remains fixed while being able to rotate. If the ball has a rough surface, the fluid flowing past it will make it spin. The ball can rotate around any axis.

We can restate the *Gradient Test* for dimension 3 as follows.

Theorem 6.10.6: Gradient Test For Dimension 3

Suppose F is a vector field on an open region in $\mathbf{R}^3$ with continuously differentiable component functions. If F is gradient (i.e., $F = \operatorname{grad} h$), then it's irrotational with respect to all three pairs of coordinates:

$$\boxed{\operatorname{rot}_{y,z} < q, r >= 0,\ \operatorname{rot}_{z,x} < r, p >= 0,\ \operatorname{rot}_{x,y} < p, q >= 0}$$

The subscripts indicate with respect to which two variables we differentiate while the third to be kept fixed.

In fact, we can form the following vector field called the *curl* of F that take care of all three rotors:

$$\operatorname{curl} F = \operatorname{rot}_{y,z} < q, r > i + \operatorname{rot}_{z,x} < r, p > j + \operatorname{rot}_{x,y} < p, q > k =< r_y - q_z,\ p_z - r_x,\ q_x - p_y > \ .$$

In particular, when the vector field $V = pi + qj + rk$ has a zero z-component, $r = 0$, while p and q don't depend on z, the curl is reduced to the rotor:

$$\operatorname{curl}(pi + qj) = \operatorname{rot} < p, q > k\,.$$

Exercise 6.10.7

Define a 4-dimensional analog of the rotor.

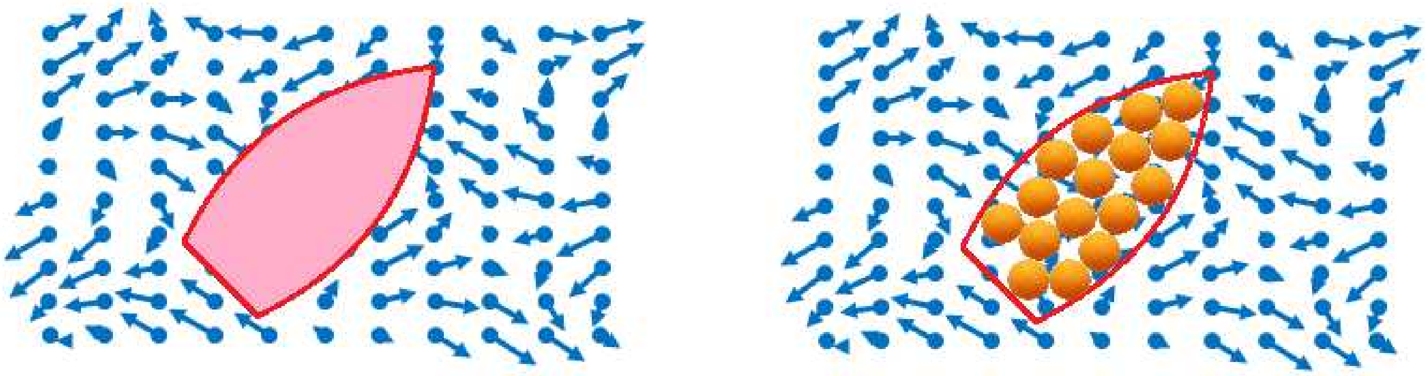
The two theorems can be restated in an even more concise form, in terms of the compositions of these *functions of functions*:

$$\operatorname{rot}\operatorname{grad} = 0 \quad \text{and} \quad \operatorname{curl}\operatorname{grad} = 0$$

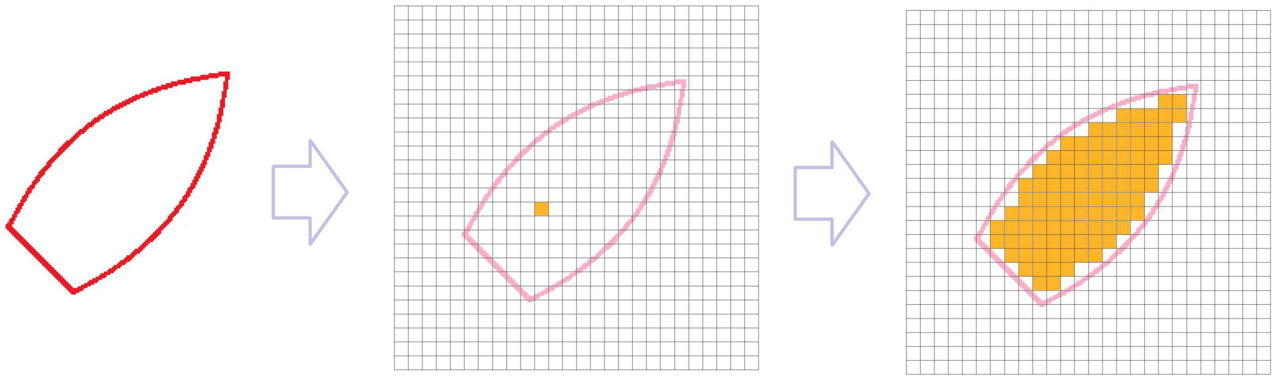
Once again, we end up at *zero* no matter what the starting point is:

$$\begin{array}{lllll}
\text{functions of two variables} & \xrightarrow{\text{grad}} & \text{vector fields in } \mathbf{R}^2 & \xrightarrow{\text{rot}} & \text{functions of two variables,} \\
\text{functions of three variables} & \xrightarrow{\text{grad}} & \text{vector fields in } \mathbf{R}^3 & \xrightarrow{\text{curl}} & \text{vector fields}
\end{array}$$

The analysis of the work integral in the continuous case is similar to the one for the discrete case. How do we measure the amount of its rotation, i.e., the work performed by the force of the flow rotating it?



We suppose that there are many little balls in the flow forming some shape and then find the amount of their total rotation, i.e., the work performed by the force of the flow rotating the balls.

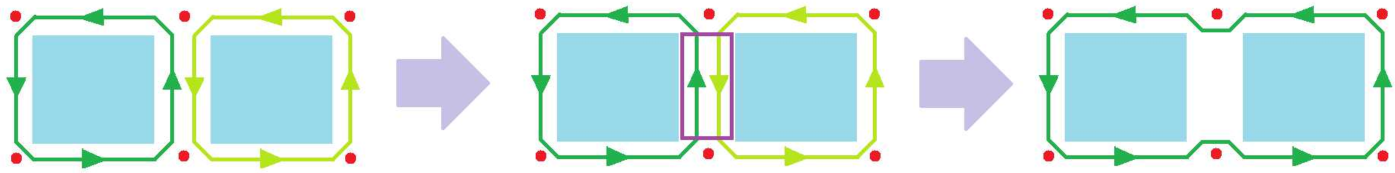


Just as before, we start with a single rectangle and then build more and more complex regions on the plane from the rectangles of our grid – as if each contains a ball – while maintaining the formula. If we have just two adjacent squares, R_1 and R_2, bounded by curves C_1 and C_2, we write Green's formula for either and

then add the two:

$$\begin{array}{rl} \displaystyle\oint_{C_1} F\cdot dX & = \displaystyle\iint_{R_1} \operatorname{rot} F\, dA \\ + & \\ \displaystyle\oint_{C_2} F\cdot dX & = \displaystyle\iint_{R_2} \operatorname{rot} F\, dA \\ \hline \displaystyle\oint_{C_1\cup C_2} F\cdot dX & = \displaystyle\iint_{R_1\cup R_2} \operatorname{rot} F\, dA \end{array}$$

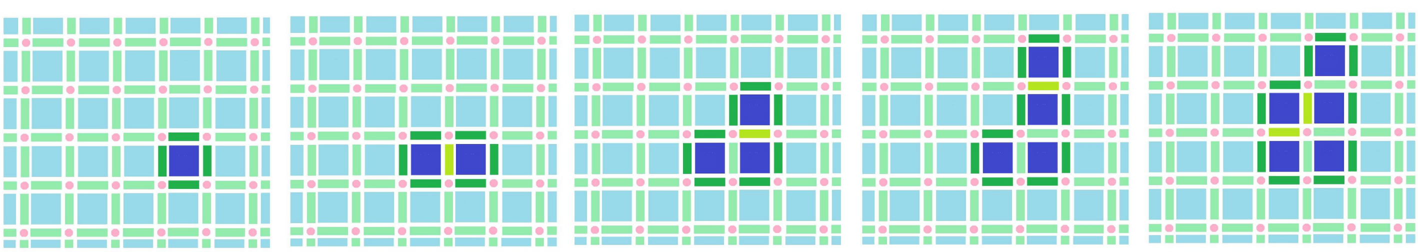
In the right-hand side, we have a single integral according to *Additivity* of double integrals and in the left-hand side, we have a single integral according to *Additivity* of line integrals. Here $C_1\cup C_2$ is the curve that consists of C_1 and C_2 traveled consecutively. The left-hand side is simplified: the two curves share an edge but travel it in the *opposite* directions.



We have a cancellation according to *Negativity* for line integrals. The result is:

$$\oint_{\partial D} F\cdot dX = \iint_D \operatorname{rot} F\, dA\,,$$

where D is the union of the two rectangles and ∂D is its boundary. We continue on adding one rectangle at a time to our region D and cancelling the edges shared with others producing bigger and bigger curve $C=\partial D$ that bounds D:



Or we can add whole regions...

It is possible, however, that the boundary curve might seize to be a single closed curve!

Theorem 6.10.8: Fundamental Theorem of Calculus for Vector Fields

Suppose a plane region R is bounded by piecewise differentiable curve C (possibly made of several disconnected pieces). Then for any vector field F with continuously differentiable components on an open set containing R, we have:

$$\boxed{\oint_C F\cdot dX = \iint_R \operatorname{rot} F\, dA}$$

Proof.

We only demonstrate the proof for a region R that has a partition that also produces a partition of C. We sample F at the secondary nodes of the partition of C and $\operatorname{rot} F$ at the tertiary nodes of the partition of R. We then use the *Fundamental Theorem of Discrete Calculus for vector fields*:

$$\sum_{\partial R} F\cdot \Delta X = \sum_R \operatorname{rot} F\,\Delta A\,.$$

We take the limits of these two Riemann sums over the partitions with the mesh approaching zero.

This is also known as *Green's Formula.* Written componentwise, it takes the following form:

$$\int_C p\,dx + q\,dy = \iint_R (q_x - p_y)\,dxdy\,.$$

Let's trace the theorem back to some familiar things.

What if the vector field is gradient? Then its rotor is zero and, therefore, our formula takes this form:

$$\oint_{\partial D} F\cdot dX = \iint_D \operatorname{rot} F\,dA = \iint_D 0\,dA = 0\,.$$

The line integral along any closed curve is then zero and, according to the *Path-independence Theorem*, F is path-independent. Then,

$$\int_C F\cdot dX = f(B) - f(A)\,,$$

for any curve C from A to B, where f is a potential function of F. We have arrived at the Fundamental Theorem of Calculus for gradient vector fields. It follows that Green's Theorem is its generalization. This confirms the role of Green's Theorem as *the* Fundamental Theorem of Calculus for all vector fields for dimension 2.

What is the vector field doesn't depend on y, i.e., $F(x,y) = F(x) =< p(x), q(x) >$, while R is a rectangle $[a,b]\times[c,d]$? First the left-hand side of the formula... The line integrals along the two horizontal sides of R cancel each other. We are left with:

$$\oint_{\partial D} F\cdot dX = F(b)\cdot A - F(a)\cdot A\,,$$

where A is the vector that represents the vertical sides of R (oriented vertically). Then,

$$\oint_{\partial D} F\cdot dX = (q(b) - q(a))(d-c)\,.$$

Now the right-hand side of the formula... The rotor is simply $q'(x)$. Then,

$$\iint_R \operatorname{rot} F\,dA = \iint_{[a,b]\times[c,d]} q'(x)\,dxdy = \int_a^b\int_c^d q'(x)\,dxdy = \int_a^b q'(x)\,dx\,(d-c)\,.$$

We have arrived to the original *Fundamental Theorem of Calculus* from Volume 3 (Chapter 3IC-1):

$$q(b) - q(a) = \int_a^b q'(x)\,dx\,.$$

Example 6.10.9: rotation vector field

Let's, again, consider this *rotation vector field*:

$$F = \frac{1}{x^2+y^2} < y,\ -x >= \left\langle \frac{y}{x^2+y^2},\ -\frac{x}{x^2+y^2} \right\rangle =< p, q >\,,$$

which is irrotational:

$$\operatorname{rot} F = q_x - p_y = 0\,.$$

Even though it passes the *Gradient Test*, it is not gradient. Indeed, if $X = X(t)$ is a counterclockwise parametrization of the circle, $F(X(t))$ is parallel to $X'(t)$, and, therefore, the line integral along the

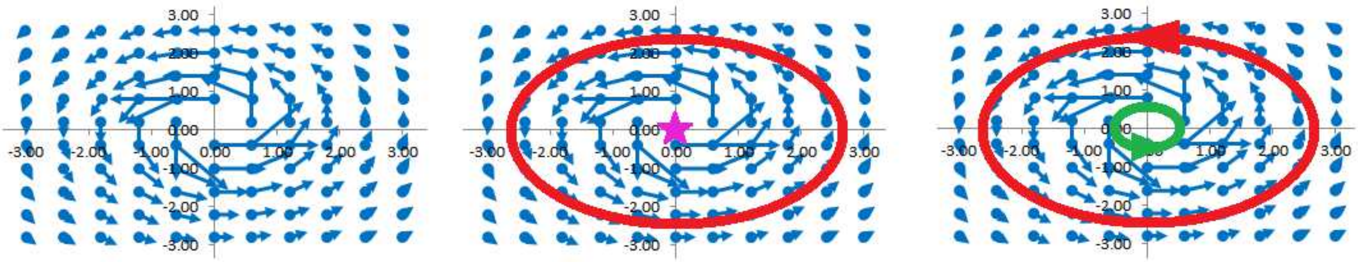
circle is positive:

$$\int_a^b F(X(t)) \cdot X'(t)\, dt > 0\,.$$

On the other hand, according to our theorem, we have:

$$\oint_C F \cdot dX = \iint_R \operatorname{rot} F\, dA = 0\,.$$

So, C doesn't bound R.



A hole is what makes a spiral staircase possible by providing a place for the pole. Now, we'd need R to be a *ring* so that the boundary of R would include another curve, maybe a smaller circle, C', going clockwise, Then,

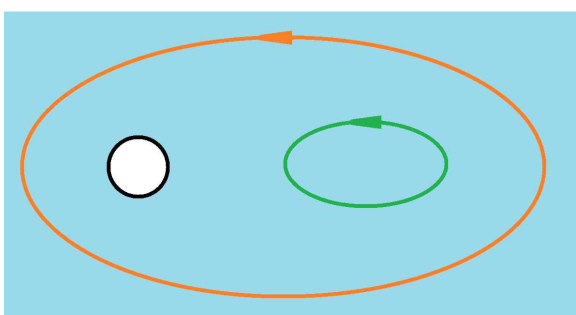
$$\oint_C F \cdot dX + \oint_{C'} F \cdot dX = \iint_R \operatorname{rot} F\, dA = 0\,.$$

Therefore, we have:

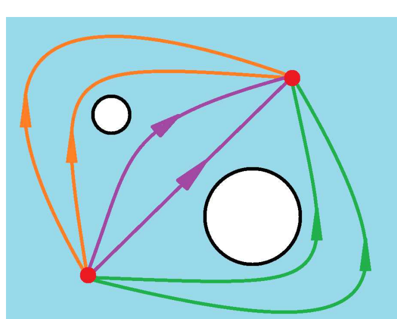
$$\oint_C F \cdot dX = \oint_{-C'} F \cdot dX\,.$$

So, moving from the larger circle to the smaller (or vice versa) doesn't change the line integral, i.e. the work. A remarkable result! It is seen as even more remarkable once we realize that the integral remains the same for all closed curves as long as they make exactly the same number of turns around the origin!

To summarize, even when the rotor is – within the region – zero, the line integral along a curve that goes around the hole may be non-zero.



Furthermore, the integral remains the same for all closed curves as long as they make exactly the same number of turns around the origin! The meaning of path-independence changes accordingly; it all depends on how the curve goes between the holes:



Example 6.10.10: walk around

Imagine that we need to find the area of a piece of land we have no access to, such as a fortification or a pond. Conveniently, Green's Formula allows us to compute area of a region without visiting the inside but by just taking a trip around it. We just need to pick an appropriate vector field:

$$F =< 0, x > \implies p = 0,\ q = x \implies p_y = 0,\ q_x = 1\,.$$

Then the formula takes the following form:

$$\begin{array}{lll}\displaystyle\iint_R (q_x - p_y)\,dxdy & = \displaystyle\int_C p\,dx & +q\,dy \\ \displaystyle\iint_R 1\,dxdy & = \displaystyle\int_C 0\,dx & +x\,dy \\ \text{Area of } R & = & \displaystyle\int_C x\,dy \end{array}$$

For example, the area of the disk R of radius r is a certain line integral around the circle C. We take C to be parametrized the usual way:

$$x = r\cos t,\ y = r\sin t\,.$$

Then,

$$\begin{aligned}\text{area of the circle } &= \int_C x\,dy \\ &= \int_0^{2\pi} r\cos t (r\sin t)'\,dt \\ &= r^2\int_0^{2\pi} \cos^2 t\,dt \\ &= r^2\big(x/2 + \sin 2x\big)\Big|_0^{2\pi} \\ &= r^2 2\pi/2 = \pi r^2\,.\end{aligned}$$

Chapter : Exercises

Contents

1. Exercises: Basic calculus

Exercise 1.1

Plot the graph of the function $y = f(x)$, where x is the income (in thousands of dollars) and $f(x)$ is the tax bill (in thousands of dollars) for the income of x, which is computed as follows: no tax on the first \$10,000, then 5% for the next \$10,000, and 10% for the rest of the income. Investigate its limits and continuity.

Exercise 1.2

Explain why the limit $\lim_{x\to 0} \sin \frac{1}{x}$ does not exist.

Exercise 1.3

(a) State the ε-δ definition of limit. (b) Use the definition to prove that $\lim_{x\to 0} x^2 = 0$.

Exercise 1.4

(a) State the definition of limit. (b) Use the definition to prove that $\lim_{x\to 0} x^3 \neq 3$.

Exercise 1.5

(a) State the definition of an infinite limit. (b) Use the definition to prove that $\lim_{x\to +\infty} x^3 = +\infty$.

Exercise 1.6

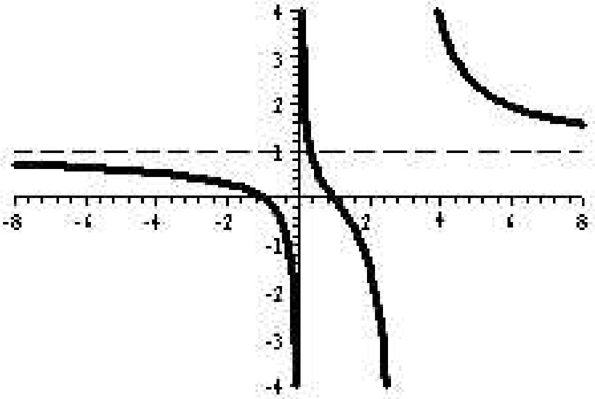
Give an example of a function with two vertical asymptotes: $x = 0$ and $x = 2$.

Exercise 1.7

Give an example of a function with a horizontal asymptote: $y = -1$, and a vertical asymptote: $x = 2$.

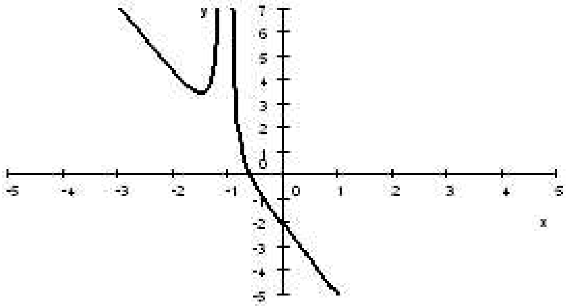
Exercise 1.8

Identify all important features of this graph:



Exercise 1.9

Express the asymptotes of this function as limits and identify other of its important features:



Exercise 1.10

True or false: "If f is continuous on (a, b), then f is bounded on (a, b)"?

Exercise 1.11

True or false: "If f is continuous on $[a, b]$, then f is bounded on $[a, b]$"?

Exercise 1.12

True or false: "If f is continuous on $[a, b)$, then f is bounded on $[a, b)$"?

Exercise 1.13

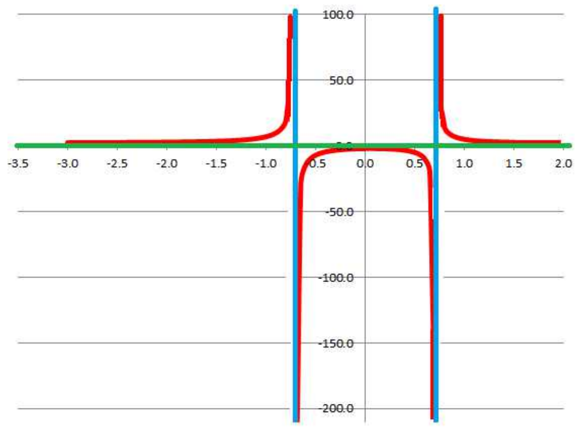
True or false: "If f is continuous on $[a, \infty)$, then f is bounded on $[a, \infty)$"?

Exercise 1.14

True or false: "Every function is bounded on a closed bounded interval"?

Exercise 1.15

The graph of f is given below. It has asymptotes. Describe them as limits. Hint: Use both $+\infty$ and $-\infty$.



Exercise 1.16

A house has 4 floors and each floor has 7 windows. What was the year when the doorman's grandmother died?

Exercise 1.17

Illustrate with plots (separately) functions with the following behavior: (a) $f(x) \to +\infty$ as $x \to 1$; (b) $f(x) \to -\infty$ as $x \to 2^+$; (c) $f(x) \to 3$ as $x \to -\infty$.

Exercise 1.18

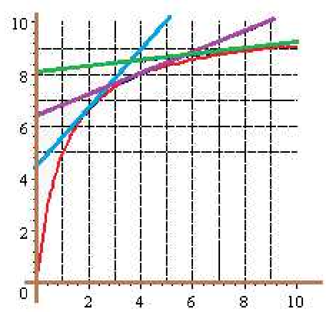
Given $f(x) = -(x - 3)^4 (x + 1)^3$. Find the leading term and use it to describe the long term behavior of the function.

Exercise 1.19

(a) State the Intermediate Value Theorem. (b) Give an example of its application.

Exercise 1.20

Three straight lines are shown below. What is so special about them? Find their slopes.



Exercise 1.21

(a) Suppose during the first 2 seconds of its flight an object progressed from point $(0, 0)$ to $(1, 0)$ to $(2, 0)$. What was its average velocity and average acceleration? (b) What if the last point is $(1, 1)$ instead?

Exercise 1.22

Suppose t is time and x is the price of bread. What can you say about its dynamics? Be as specific as possible.

t	x	y	z
0.00	1.00	1.00	4.00
0.20	0.83	1.36	4.03
0.40	0.71	1.64	4.07
0.60	0.63	1.84	4.09
0.80	0.56	1.96	4.07
1.00	0.50	2.00	4.00
1.20	0.45	1.96	3.87
1.40	0.42	1.84	3.67
1.60	0.38	1.64	3.41
1.80	0.36	1.36	3.07
2.00	0.33	1.00	2.67
2.20	0.31	0.56	2.19
2.40	0.29	0.04	1.63

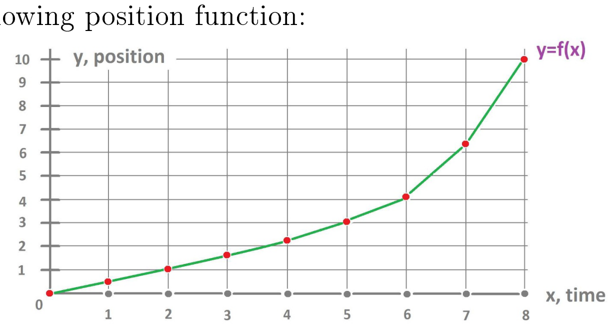
Exercise 1.23

Find the difference quotients for the function given by the following data:

x	$y = f(x)$
-1	2
1	2
3	3
5	3
7	-2
9	5

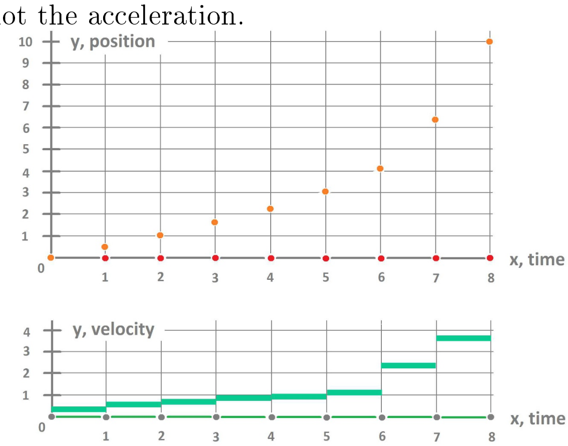
Exercise 1.24

Plot the graph of the average velocity for the following position function:



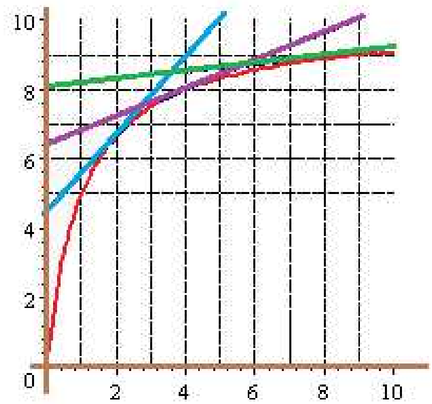
Exercise 1.25

The position and the velocity are plotted below. Plot the acceleration.



Exercise 1.26

Each of these straight lines are drown through two point of the graph. What do they tell us about the function?



Exercise 1.27

From the definition, compute the average rate of change for the function $f(x) = x^2 + 1$ at $a = 2$ with $h = 0.2$ and $h = 0.1$. Explain the difference.

Exercise 1.28

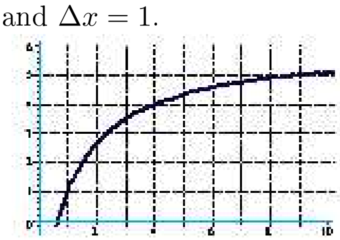
(a) Compute the average rate of change for the function $f(x) = 3x^2 - x$ at $a = 1$ and $h = .5$.
(b) Find the equation of the secant to the graph of $y = f(x)$ corresponding to this average rate of change.

Exercise 1.29

The graph of a function $f(x)$ is given below. Estimate the values of the difference quotient $\frac{\Delta f}{\Delta x}$ for $x = 0, 4$, and 6 and $\Delta x = .5$.

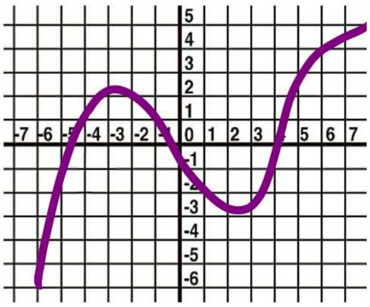
Exercise 1.30

The graph of a function $f(x)$ is given below. Estimate the values of the difference quotient for $x = 2, 4, 9$ and $\Delta x = 1$.



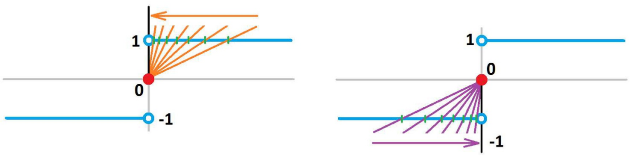
Exercise 1.31

The graph of a function f is given below. Estimate the values of the difference quotient $\frac{\Delta f}{\Delta x}$ for $x = 1$ and $\Delta x = 2,\ 1,\ .5$.



Exercise 1.32

The secant line of the sign function are shown below. What do they tell you about the differentiability of the function at $x = 0$?



Exercise 1.33

You have received the following email from your boss: "Tim, Look at the numbers in this spreadsheet. This stock seems to be inching up... Does it? If it does, how fast? Thanks. – Tom". Describe your actions.

Exercise 1.34

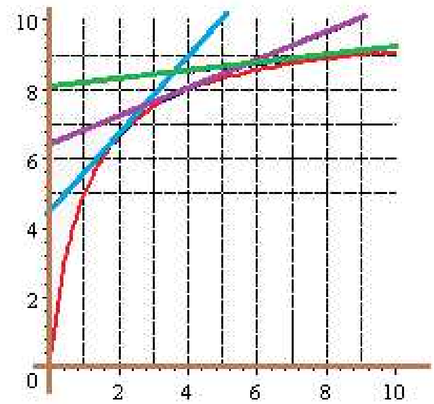
If two functions are equal, do their derivatives have to be equal too?

Exercise 1.35

Find the tangent line through the point $(2, 1)$ to the graph of the function the derivative of which is e^{x^2}.

Exercise 1.36

What do these straight lines tell us about the function?



Exercise 1.37

(a) State the definition of the derivative of a function at point a. (b) Provide a graphical interpretation of the definition.

Exercise 1.38

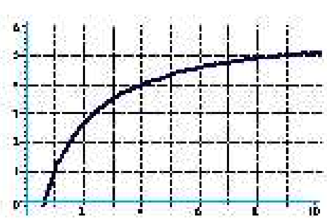
From the definition, compute the derivative of $f(x) = x^2 + 1$ at $a = 2$.

Exercise 1.39

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 0, 4,$ and 6.

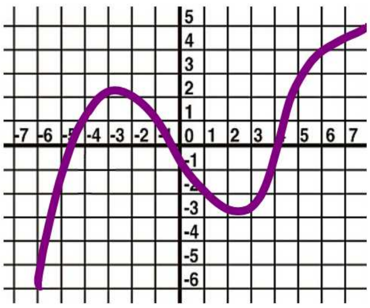
Exercise 1.40

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 2, 4,$ and 9.



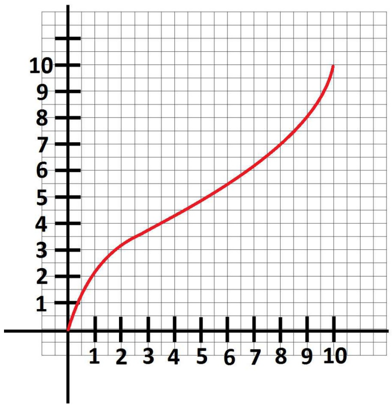
Exercise 1.41

The graph of a function f is given below. Estimate the values of the derivative f' for $x = 0$ and $x = 4$. Show your computations.



Exercise 1.42

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 1, 3$, and 6.



Exercise 1.43

Find all local maxima and minima of the function $f(x) = x^3 - 3x - 1$.

Exercise 1.44

(a) Analyze the first and second derivatives of the function $f(x) = x^4 - 2x^2$. (b) Use part (a) to sketch its graph of f.

Exercise 1.45

Suppose the functions that follow are differentiable. (a) Finish the statement "If $h'(x) = 0$ for all x in (a, b), then...". (b) Finish the statement "If $f'(x) = g'(x)$ for all x in (a, b), then...".

Exercise 1.46

Sketch the graph of the function $f(x) = \sqrt{x}e^{-x}$. Justify the graph by studying the derivatives of f.

Exercise 1.47

(1) State Rolle's Theorem and illustrate it with a sketch. (b) Quote and state the theorem(s) necessary to prove it. (c) What theorem follows from it?

Exercise 1.48

Sketch the graph of the function given below. Provide justification for each feature of the graph:

$$f(x) = \frac{x^2 + 7x + 3}{x}.$$

Exercise 1.49

(a) State the Mean Value Theorem. (b) Verify that the function $f(x) = \frac{x}{x+2}$ satisfies the hypotheses of the theorem on the interval $[1, 4]$.

Exercise 1.50

Sketch the graph of the function $f(x) = x^4 - x^2$. Provide justification for each feature of the graph.

Exercise 1.51

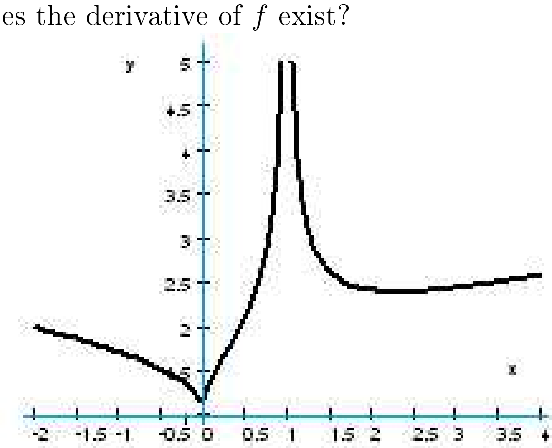
Find global maxima and minima of the function, $f(x) = x^3 - 3x$ on the interval $[-2, 10]$.

Exercise 1.52

Find the local maximum and minimum points of the function $f(x) = x^3 - 3x$.

Exercise 1.53

The graph of function f is given below. (a) At what points is f continuous? (b) At what points does the derivative of f exist?



Exercise 1.54

Indicate which the following statements below is true or false (no proof necessary):

1. If the function f is increasing, then so is f^{-1}.
2. The exponential function has an asymptote.

3. If $f'(c) = 0$, then c is a local maximum or a local minimum of f.
4. If a function is differentiable then it is continuous.
5. If two functions are equal, their derivatives are also equal.
6. If two functions are equal, their antiderivatives are also equal.

Exercise 1.55

(a) State the Fundamental Theorem of Calculus. (b) Use part (a) to evaluate

$$\int_{-1}^{1} \sin\frac{x}{3}\,dx\,.$$

Exercise 1.56

(a) State the Fundamental Theorem of Calculus. (b) Use part (a) to evaluate

$$\frac{d}{dx}\int_0^x e^{t^2}\,dt\,.$$

Exercise 1.57

(a) Make a sketch of the left-end Riemann sums for $\int_0^1 \sqrt{x}\,dx$ with $n = 4$ intervals. (b) State the algebraic properties of the Riemann integral.

Exercise 1.58

Given $f(x) = x^2 + 1$, write (but do not evaluate) the Riemann sum for the integral of f from -1 to 2 with $n = 6$ and left ends as sample points. Make a sketch.

Exercise 1.59

Provide the definition of the definite integral via its Riemann sums. Make a sketch.

Exercise 1.60

The Fundamental Theorem of Calculus includes the formula $\int_a^b f(x)\,dx = F(b) - F(a)$. (a) State the whole theorem. (b) Provide definitions of the items appearing in the formula.

Exercise 1.61

(a) State the definition of the definite integral $\int_a^b f(x)\,dx$ and illustrate the construction with a sketch. (b) Use the definition to justify that

$$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$$

for a constant c.

Exercise 1.62

Suppose

$$\int_0^1 f\,dx = 2,\ \int_0^4 f\,dx = 0,\ \int_1^2 f\,dx = 2\,.$$

Find

$$\int_1^3 f\,dx,\ \int_0^1 (f(x)+3)\,dx,\ \int_2^4 f\,dx\,.$$

Exercise 1.63

Suppose a function is defined by:

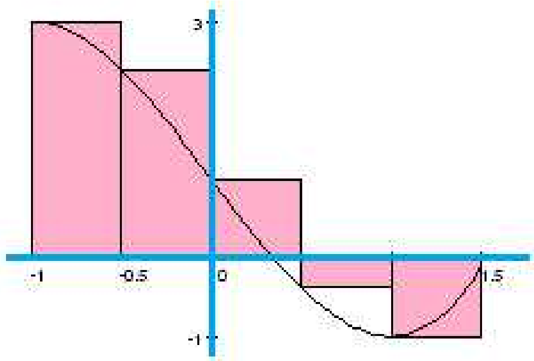
$$F(x) = \int_2^x f\,dx\,.$$

Find, in terms of F, the following:

$$\int_0^4 f\,dx,\ \int_1^2 f\,dx,\ \int_0^{-1} f\,dx,\ \int_1^2 (f(x)-1)\,dx\,.$$

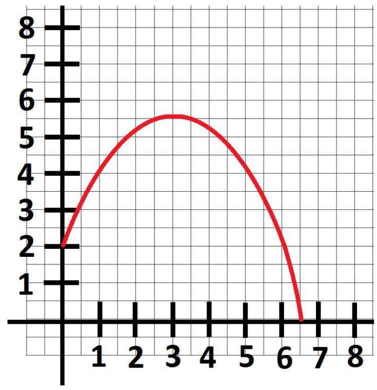
Exercise 1.64

Evaluate the Riemann sum of f below on the interval $[-1, 1.5]$ with $n = 5$. What are its sample points? What does it estimate?



Exercise 1.65

Write (don't evaluate) the left-end Riemann sum of the integral $\int_0^5 f(x)\,dx$ for function f shown below with $n = 5$ intervals.

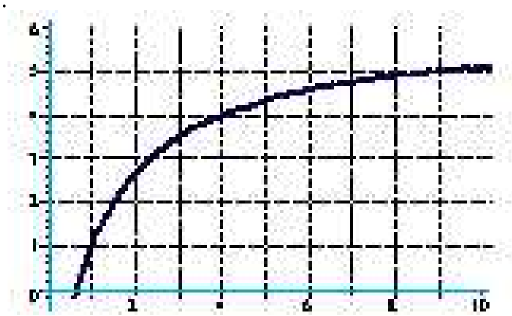


Exercise 1.66

Write the mid-point Riemann sum that approximates the integral $\int_0^1 \sin x\, dx$ within .01.

Exercise 1.67

Let $I = \int_2^8 f\, dx$. (a) Use the graph of $y = f(x)$ below to estimate L_4, M_4, R_4. (b) Compare them to I.



Exercise 1.68

Complete the following statements:

- $(f(x) \cdot x^2)' = f'(x) \cdot x^2 + \ ...$
- $\int x^{-1}\, dx = \ ...$
- $\int f'(x)\, dx = \ ...$
- $\int u\, dv = uv\ ...$
- $u = \cos t \implies du = \ ...$

Exercise 1.69

Suppose that F is an antiderivative of a differentiable function f. If F is increasing on $[a, b]$, what can you say about f?

Exercise 1.70

Execute the following substitution in the integral (don't evaluate the resulting integral):

$$\int \sqrt{\cos x + \sin x}\, dx, \quad u = \sin x\,.$$

Exercise 1.71

Suppose $s(t)$ represents the position of a particle at time t and $v(t)$ its velocity. If $v(t) = \sin t - \cos t$ and the initial position is $s(0) = 0$, find the position $s(1)$.

Exercise 1.72

Evaluate

$$\int e^{3x}\, dx\,.$$

Exercise 1.73

Evaluate

$$\int e^{x^2} 2x\, dx\,.$$

Exercise 1.74

Evaluate

$$\int 2x \sin 5x\, dx\,.$$

Exercise 1.75

Evaluate

$$\int_1^3 e^{t+1}\, dx\,.$$

Hint: Watch the variables.

Exercise 1.76

Calculate:

$$\int \left(e^{\sin x^2 + 77}\right)'\, dx\,.$$

Exercise 1.77

Evaluate the integral by substitution

$$\int x e^{x^2}\, dx\,.$$

Exercise 1.78

Find all antiderivatives of the following function: $f(x) = e^{-x}$.

Exercise 1.79

Find the antiderivative F of the function $f(x) = 3x^2 - 1$ satisfying the initial condition $F(1) = 0$.

Exercise 1.80

Evaluate the integral

$$\int_0^1 x^3\,dx\,.$$

Exercise 1.81

Evaluate:

$$\int x^2\,dx - \int x^2\,dx\,.$$

Exercise 1.82

Integrate by parts:

$$\int 3xe^{-x}\,dx\,.$$

Exercise 1.83

Use the table of integrals to evaluate:

$$\int \sin^{-1} 2x\,dx\,.$$

Exercise 1.84

Evaluate:

$$\int_0^1 \frac{1}{2x}\,dx\,.$$

Exercise 1.85

Use substitution to evaluate the integral:

$$\int_0^\pi \sin x\ \cos^2 x\,dx\,.$$

Exercise 1.86

Use substitution to evaluate the integral:

$$\int_0^1 \frac{1}{\sqrt{4-x^2}}\,dx\,.$$

Exercise 1.87

Use the table of integrals to evaluate:

$$\int x^2(\sqrt{x^2-4}-\sqrt{x^2+9})\,dx\,.$$

Exercise 1.88

Use substitution $u = 1+x^2$ to evaluate the integral

$$\int \sqrt{1+x^2}x^5\,dx\,.$$

Exercise 1.89

Use substitution to evaluate the integral:

$$\int_0^\pi \sin x\ \cos^2 x\,dx\,.$$

Exercise 1.90

Evaluate the improper integral:

$$\int_1^\infty \frac{1}{2x}\,dx\,.$$

Exercise 1.91

Find the antiderivative F of the function $f(x) = e^x + x$ satisfying the initial condition $F(0) = 1$.

Exercise 1.92

The region bounded by the graphs of $y = \sqrt{x}$, $y = 0$, and $x = 1$ is revolved about the x-axis. Find the surface area of the solid generated.

Exercise 1.93

A chord of a circle is a straight line segment whose end-points lie on the circle. Find the average length of a chord perpendicular to the diameter. What about parallel?

Exercise 1.94

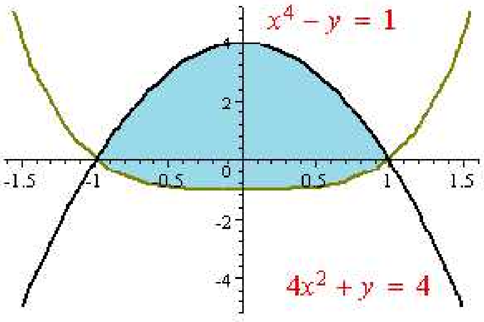
Find the average length of a segment in a square parallel to (a) the base, (b) the diagonal.

Exercise 1.95

Find (by integration) the length of a circle of radius r.

Exercise 1.96

Find the area enclosed by the curves below:



Exercise 1.97

Find the area of the region bounded by $y = x^2 - 1$ and $y = 3$.

Exercise 1.98

Suppose f is an integrable function. (a) Show that f is also odd then $\int_{-a}^{a} f\,dx = 0$. (b) Suggest a related formula for an even f.

Exercise 1.99

Find the centroid of the region bounded by the curves $y = x^2,\ y = 1$.

Exercise 1.100

Find the x-coordinate of the center of mass of the region between $y = x^2$ and $y = x^3$.

Exercise 1.101

Find the volume of a right circular cone of radius R and height h by any method you like.

Exercise 1.102

Compute the average area of the cross section of the sphere of radius 1.

Exercise 1.103

Find the center of mass of the region below $y = 2x$ for $0 \le x \le 1$.

Exercise 1.104

The volume of a solid is the integral of the areas of its cross-sections. Explain and justify using Riemann sums.

Exercise 1.105

The region bounded by the graphs of $y = x^2 + 1,\ y = 0,\ x = 0$ and $x = 1$ is revolved about the x-axis. Find the volume area of the solid generated.

Exercise 1.106

The region bounded by the graphs of $y = x^2 + 1,\ y = 0,\ x = 0$, and $x = 1$ is revolved about the y-axis. Find the volume area of the solid generated.

Exercise 1.107

An aquarium 2 m long, 1 m wide, and 1 m deep is full of water. Find the work needed to pump half of the water out of the aquarium (the density of water is 1000 kg/m^3).

Exercise 1.108

Find the area of the surface of revolution around the x-axis obtained from $y = \sqrt{x},\ 4 \le x \le 9$.

Exercise 1.109

Find the centroid of the region bounded by the curves $y = 4 - x^2,\ y = x + 2$.

Exercise 1.110

Find the area of the region bounded by $y = x^2 - 1$ and $y = 3$.

Exercise 1.111

Find the area under the graph of the function $f(x) = e^x$ from $x = -1$ to $x = 1$.

Exercise 1.112

Find the average value of the function $f(x) = 2x^2 - 3$ on the interval $[1, 3]$.

Exercise 1.113

Find the area of the region bounded by $y = \sqrt{x}$, the x-axis, and the lines $x = 1$ and $x = 4$.

2. Exercises: Algebra and geometry

Exercise 2.1

Set up, but do not solve, a system of linear equations for the following problem: "Suppose your portfolio is worth $\$1,000,000$ and it consists of two stocks A and B. The stocks are priced as follows: A $\$2.1$ per share, B $\$1.5$ per share. Suppose also that you have twice as much of stock A than B. How much of each do you have?"

Exercise 2.2

Give the number t that makes $X =< 3, 2, 1 >$ and $Y =< 2, t, t >$ perpendicular.

Exercise 2.3

Here are xyz-equations for two planes: $x+y-z = 0$ and $x - y + z = 0$. Explain how you can tell that these planes cut each other NOT at right angles.

Exercise 2.4

A plane has an xyz-equation $x + y = 2$. Give a vector perpendicular to the plane.

Exercise 2.5

In an effort to find the line in which the planes $2x - y - z = 2$ and $-4x + 2y + 2z = 1$ intersect, a student multiplied the first one by 2 and then added the result to the second. He got $0 = 5$. Explain the result.

Exercise 2.6

Determine whether these points lie on a straight line:

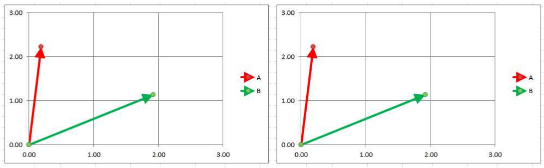
$$A = (0, -5, 5),\ B = (1, -2, 4),\ C = (3, 4, 2)\,.$$

Exercise 2.7

Find the plane through the point $P = (-1, 6, -5)$ and parallel to the vectors $A =< 1, 1, 0 >$ and $B =< 0, 1, 1 >$.

Exercise 2.8

Vectors A and B are given below. Copy the picture and illustrate graphically (a) $A+B$, (b) $A-B$, (c) $||A||$, (d) the projection of A on B, (e) the projection of B on A.



Exercise 2.9

Find the angle between the vectors $< 1, 1, 1 >$ and the x-axis. Don't simplify.

Exercise 2.10

Find the plane through the origin perpendicular to the line from $(1, 0, 0)$ and $(0, 1, 1)$.

Exercise 2.11

Find an equation of the plane through $(2, 1, 0)$ and parallel to $x + 4y - 3z = 1$.

Exercise 2.12

(a) Find the angle between the planes $x+y+z = 1$ and $x - 2y + 3z = 1$. (b) Find the equation of the line of intersection of these planes.

Exercise 2.13

Find the vector equation of the line parallel to both xy- and xz- coordinate planes an passing through $(2, 3, 1)$.

Exercise 2.14

Solve the system of linear equations:

$$\begin{cases} x & -y & = -1\,, \\ 2x & +y & = 0\,. \end{cases}$$

Exercise 2.15

Find the reduced row echelon form of the following system of linear equations. What kind of set is its solution set?

$$\begin{cases} -x & -2y & +z & = 0\,, \\ 3x & & +z & = 2\,, \\ x & -y & +z & = 1\,. \end{cases}$$

Exercise 2.16

Represent the system of linear equations as a matrix equation:

$$\begin{cases} x & -y & +z & = -1\,, \\ 3x & & +z & = 2\,, \\ 2x & +y & +z & = 1\,. \end{cases}$$

Exercise 2.17

Represent this matrix equation as a system of linear equations:

$$\begin{bmatrix} 1 & 3 & 2 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Exercise 2.18

Give explicitly the solution set of the system of linear equations represented by its augmented matrix:

$$\begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 1 & 2 \end{bmatrix}.$$

Exercise 2.19

Find scalars a and b such that

$$a < 1,2 > +b < -1,3 >=< 1,7 > \ .$$

Exercise 2.20

Is it possible for a system of linear equations to have: (a) no solutions, (b) exactly one solution, (c) exactly two solutions, (d) infinitely many solutions? Give an example or explain why it's not possible.

Exercise 2.21

Find the set of all vectors in $\mathbf{R}^2$ that are orthogonal to $< -1,3 >$. Write the set in the standard form of a line through the origin.

Exercise 2.22

Compute:

$$\begin{bmatrix} 1 & 3 & 2 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

3. Exercises: Parametric curves

Exercise 3.1

Describe the motion of a particle with position (x, y), where

$$x = 2 + t\cos t,\ \ y = 1 + t\sin t\,,$$

as t varies within $[0, \infty)$.

Exercise 3.2

Suppose the parametric curve is given by

$$x = \cos 3t,\ \ y = 2\sin t\,.$$

Set up, but do not evaluate, the integrals that represent (a) the arc-length of the curve, (b) the area of the surface obtained by rotating the curve about the x-axis.

Exercise 3.3

Suppose curve C is the graph of function $y = f(x)$. (a) Find a parametric representation of C. (b) Find a parametric representation of C that goes from right to left.

Exercise 3.4

Find all points on the curve

$$x = \cos 3t,\ \ y = 2\sin t\,,$$

where the tangent is either horizontal or vertical.

Exercise 3.5

Sketch the following parametric curve:

$$x = |\cos t|,\ \ y = |\sin t|,\ \ -\infty < t < +\infty\,.$$

Describe the curve and the motion.

Exercise 3.6

Sketch the following parametric curves:

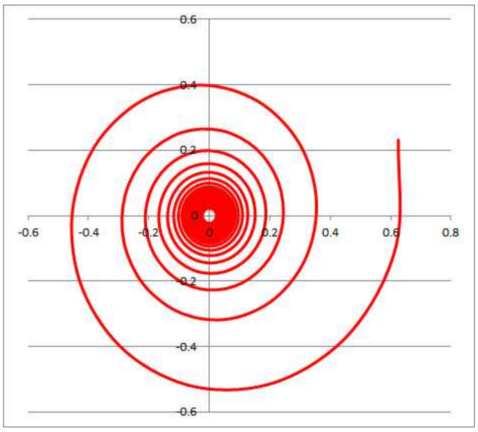
- $x(t) = \dfrac{1}{t}\,,\ \ y(t) = \sin t\,,\ \ t > 0$
- $x = \cos t\,,\ \ y = 2$
- $x = 1/t,\ \ y = 1/t^2\,,\ \ t > 0$

Exercise 3.7

(1) Sketch the parametric curve $x = \cos t,\ y = \sin 2t$. (2) The curve intersects itself. Find the angle of this intersection.

Exercise 3.8

Find an equation of the spiral converging to the origin as below:

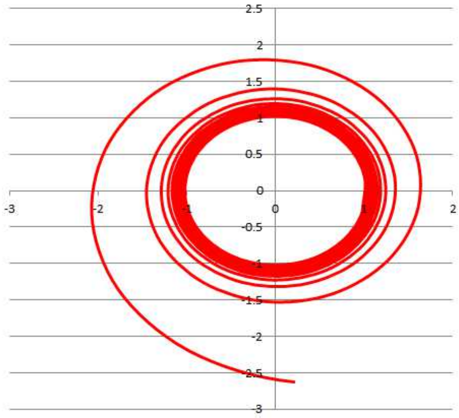


Exercise 3.9

Plot this entire parametric curve: $x = \sin t,\ y = \cos 2t$.

Exercise 3.10

Find a parametric representation of a curve similar to the one below, a spiral wrapping around a circle. What about one that is wrapping from the inside? (no proof necessary):



Exercise 3.11

Given a parametric curve $x = \sin t,\ y = t^2$. Find the line(s) tangent to the curve at the origin.

Exercise 3.12

Find a parametric representation of a curve that looks like the figure eight or a flower (no proof necessary).

Exercise 3.13

Sketch the parametric curve:

$$x = t^2 - 1,\ y = 2t^2 + 3\,.$$

Exercise 3.14

Suppose that the following parametric curve represents the motion of an object on the plane:

$$x = 3t - 1,\ y = t^2 - 1\,.$$

(a) When does the object cross the x-axis? (b) When does the object cross the y-axis?

Exercise 3.15

Represent as a parametric curve the rotation of a rod of length 2 that makes one full turn every 3 seconds.

Exercise 3.16

One circle is centered at $(0, 0)$ and has radius 1. The second is centered at $(3, 3)$. What is the radius of the second if the two circles touch?

Exercise 3.17

Represent in polar coordinates these points given by their Cartesian coordinates: (a) $(1, 2)$; (b) $(-1, -1)$; (c) $(0, 0)$.

Exercise 3.18

Represent in Cartesian coordinates these points given by their polar coordinates: (a) $\theta = 0,\ r = -1$; (b) $\theta = \pi/4,\ r = 2$; (c) $\theta = 1,\ r = 0$.

Exercise 3.19

(a) Represent the following complex number in the standard form: $(2+3i)(-1+2i)$. Indicate the real and imaginary parts. (b) Find its module and argument.

Exercise 3.20

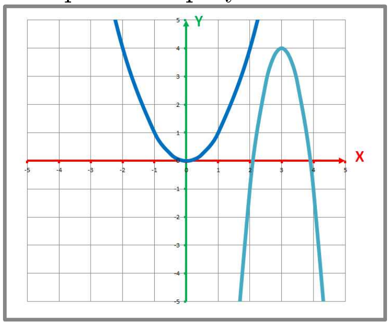
Simplify $(1 + i)^2$.

Exercise 3.21

(a) Find the roots of the polynomial $x^2 + 2x + 2$. (b) Find its x-intercepts. (c) Find its factors.

Exercise 3.22

What can you say about the imaginary parts of the roots of these quadratic polynomials?



Exercise 3.23

Suppose $h(t) = (\sin e^t, \cos e^t)$ is a parametric curve. (a) What is its path? (b) Show how the Chain Rule is used to compute its derivative.

Exercise 3.24

Suppose you are towing a trailer-home. During the first few minutes, every time you look at the rear view mirror you can see only the lower part of the home. Later, every time you look you can see only the top part. Discuss the profile of the road.

Exercise 3.25

Suppose during the first 2 seconds of its flight an object progressed from point $(0, 0)$ to $(1, 0)$ to $(1, 1)$. What was its (a) average velocity, (b) average acceleration?

Exercise 3.26

Sketch the following parametric curve:

1. $x(t) = \dfrac{1}{t}\,, y(t) = \sin t,\ t > 0$
2. $< x\,, y > = < \cos t\,, 2 >$
3. $r = 2^t i + 2^{2t} j$
4. $x = 1/t\,, y = 1/t^2\,,\ t > 0$

Exercise 3.27

Find all points on the curve

$$x = t\cos t\,,\ y = t\sin t\,,\ t > 0\,,$$

where the tangent vector points in the following direction: 1. up, 2. left, 3. down, 4. right.

Exercise 3.28

Suppose a ball in thrown horizontally at speed v feet per second by a person h feet tall. Represent the motion as a parametric curve in the 3-

dimensional space.

Exercise 3.29

Sketch the parametric curve

$$x = \frac{1}{t}\cos t\,,\ y = \frac{1}{t}\sin t\,,\ t > 0\,.$$

Exercise 3.30

Suppose an object is dropped from a 100 feet building. You are standing 100 feet away from the building and tracing the object with a laser. Express the angle of the laser as a function of time.

Exercise 3.31

At time t with $0 \leq t$, an object is at the position

$$P(t) = \left(t^2 + 1\,, \frac{1}{t+1}, e^{2t}\right)\,.$$

Calculate its velocity, $V(t)$, and its acceleration, $A(t)$, as functions of t.

Exercise 3.32

Parametrically describe the line segment with endpoints $(-1,-1,-1)$ and $(1,1,1)$.

Exercise 3.33

"If f is continuous at x_0 then

$$\lim_{x\to x_0} \frac{f(x)}{e^x}$$

exists." True or false? Explain.

Exercise 3.34

Find a parametric curve such that $F'(t) = (e^t\,, \sin t)$ and $F(0) = (0,1)$.

Exercise 3.35

Show that for any parametric curve $\lim\limits_{t\to t_0} \frac{1}{t}F(t) = 0$ implies $\lim\limits_{t\to t_0} F(t) = 0$.

Exercise 3.36

Let $X = F(t)$ be a differentiable parametric curve. If $F'(t)$ is perpendicular to $F(t)$ for all t, show that $||F(t)||$ is constant.

Exercise 3.37

(1) Sketch the parametric curve $x = \cos t\,,\ y = \sin 2t$. (2) The curve intersects itself. Find the angle of this intersection.

Exercise 3.38

If the velocity of an object at time t is given by $V(t) =< 1 + t^2\,, \sqrt{t} >$, what is its position at time $t = 3$ if the objects starts at the origin?

Exercise 3.39

Find the unit tangent vector of the curve

$$F(t) = ti + tj + (1 + t^2)k$$

at the point $(0,0,1)$.

Exercise 3.40

Plot and describe the curve

$$x(t) = |\sin t|\,,\ y(t) = |\cos t|\,,\ z = t\,.$$

Exercise 3.41

Find the line tangent to the curve

$$F(t) = (t^5\,, t^4\,, t^3)$$

at point $(1,1,1)$.

Exercise 3.42

Find the unit normal vector of the curve

$$G(t) = 3ti + 2t^2 j\,.$$

Exercise 3.43

Find the curvature of the curve $< t, t^2 >$.

Exercise 3.44

Find the equation of the line tangent to the curve

$$x = t^5\,,\ y = t^4\,,\ z = t^3\,,$$

at the point $(1,1,1)$.

Exercise 3.45

Sketch the following parametric curve:

$$x = |\cos t|\,,\ y = |\sin t|\,,\ -\infty < t < +\infty\,.$$

Describe the curve and the motion.

Exercise 3.46

Find the line tangent to the curve

$$F(t) = < t^5 \, , \, t^4 \, , \, t^3 >$$

at the point $(1, 1, 1)$.

Exercise 3.47

Suppose a ball in thrown horizontally at speed w feet per second by a person h feet tall. At what speed will the ball hit the ground?

Exercise 3.48

Give the definition of the curvature. Give examples of curves with various curvatures.

Exercise 3.49

Find all points on the curve $x = 3 \cos t$, $y = 2 \sin t$ where the tangent vector points in the following direction: (a) up, (b) left, (c) down, (d) right.

Exercise 3.50

Sketch a curve on the plane the curvature of which is: (a) increasing, (b) decreasing, (c) constant non-zero, (d) zero.

Exercise 3.51

Sketch the following parametric curve:

$$x = t^3 \, , \; y = t^5 \, , \; -1 \leq t \leq 1 \, .$$

Is this a regular parametrization?

Exercise 3.52

(a) Give the definition of the curvature (as a certain derivative). (b) Use the definition to compute the curvature of a circle of radius R.

Exercise 3.53

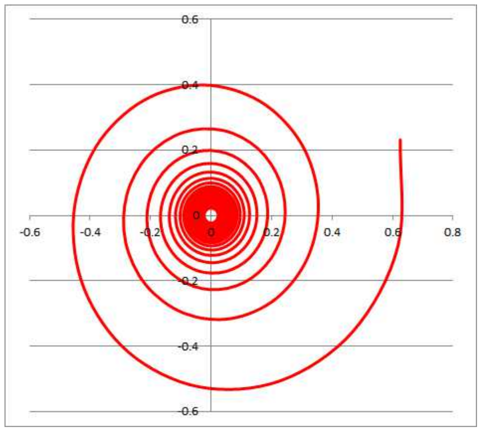
Find the line tangent to the curve

$$r(t) = < t^5, t^4, t^3 >$$

at the point $(1, 1, 1)$.

Exercise 3.54

Find an equation of the spiral in space converging to the origin as below (view from above):



Exercise 3.55

An artillery gun with a muzzle velocity of 1000 ft/s is located atop a seaside cliff 500 ft high. At what initial inclination angle should it fire a projectile in order to hit a ship at sea 20,000 ft from the foot of the cliff? Assume $g = 32 \; ft/s^2$.

Exercise 3.56

Find the center of curvature of the parabola $y = x^2$ at the point $(1, 1)$.

Exercise 3.57

Find the arc-length of the curve $x = 2e^t$, $y = e^{-t}$, $z = 2t$, from $t = 0$ to $t = 1$.

Exercise 3.58

Describe and sketch the parametric curve $f(t) = (t^3 - t, 1 - t^2)$.

Exercise 3.59

Represent as a parametric curve with domain $(-\infty, \infty)$ a plan spiral approaching 0 but never reaching it.

Exercise 3.60

Find the tangent line to the curve $f(t) = (t, t^2, t^3)$ at the point $(1, 1, 1)$.

Exercise 3.61

Show that the parametrization of the curve $f(t) = (3 \sin t \, , 4 \cos t)$ isn't natural. Find the equation for the natural parameter. Do not solve.

Exercise 3.62

Find the function $f : \mathbf{R} \to \mathbf{R}^2$ such that $f''(t) = (\cos t \, , \sin 3t)$ and $f(0) = (1, 0)$, $f'(0) = (0, 0)$.

Exercise 3.63

Let $f(t) = (t^3 - 3t,\, t^2)$. (a) Find the derivative f' of f. (b) Use f' to plot the parametric curve f.

Exercise 3.64

Use the arc-length formula to compute the length of the circle of radius 3 centered at $(1, 1)$.

Exercise 3.65

(a) Find the natural parametrization of the helix $F(t) = (\cos t\,,\, \sin t,\; t)$. (b) Find its curvature. (c) Find the center of the osculating circle at $(1, 0, 0)$.

Exercise 3.66

The curve $(t^3 - t\,,\, t^2 - 1)$ has a self-intersection at 0. Compute the angle.

Exercise 3.67

Here are parametric equations for two lines: $(x, y, z) = (1 + 2t, 3t, 2 + 5t)$ and $(x, y, z) = (-1, -1, -1) + t < 1, 0, 1 >$. Say how you can tell that these lines are NOT parallel.

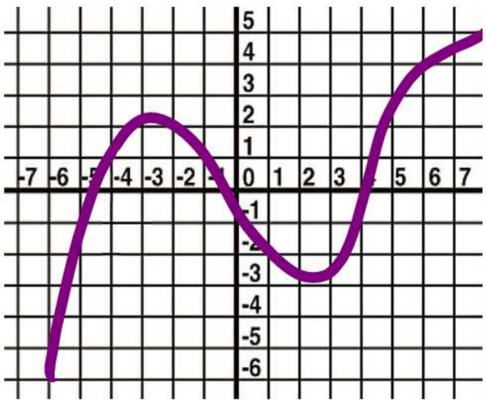
4. Exercises: Functions of several variables

Exercise 4.1

Draw a few level curves of the function $f(x,y) = x^2 + y$.

Exercise 4.2

The graph of function $y = g(x)$ of *one* variable is shown below. Suppose now that $z = f(x,y) = g(x)$ is a function of *two* variables, which depends only on x, given by the same formula. Find all points where the gradient of f is equal to 0.

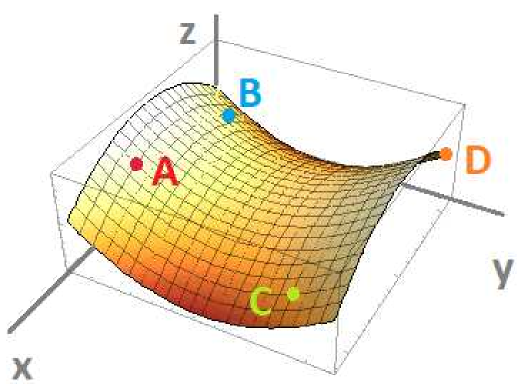


Exercise 4.3

Find all critical points of the function $f(x,y) = 2x^3 - 6x + y^2 - 2y + 7$.

Exercise 4.4

Sketch the contour (level) curves of the function shown below, along with points A, B, C, D, on the xy-plane:



Exercise 4.5

Sketch the level curves of the function $f(x,y) = 2xy + 1$ for the following values of $z = -1, 0, 1, 2$.

Exercise 4.6

Show that the limit doesn't exist:

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2 + y^2}\,.$$

Exercise 4.7

Draw the contour map (level curves) of the function $f(x,y) = e^{y/x}$. Explain what the level curves are.

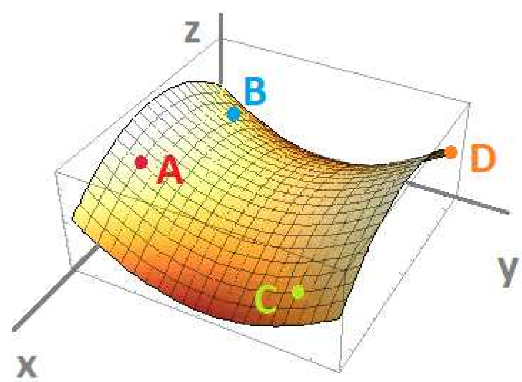
Exercise 4.8

Sketch the graph of a function of two variables $z = f(x,y)$ the derivatives of which have the following signs:

$$f_x > 0\,,\ f_{xx} > 0\,,\ f_y < 0\,,\ f_{yy} < 0\,.$$

Exercise 4.9

The graph of a function of two variables $z = f(x,y)$ is given below along with four points on the graph. Sketch the gradient for each on a separate xy-plane:

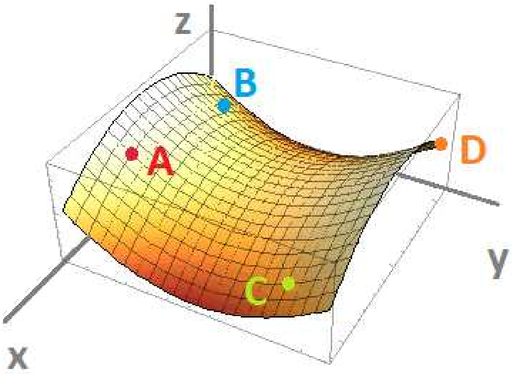


Exercise 4.10

Find the gradient of the function $f(x,y) = x^2 y^{-3}$ at the point $(1,1)$. Use this information to sketch the graph of f in the vicinity of this point. Explain.

Exercise 4.11

The graph of a function of two variables $z = f(x,y)$ is given below along with four points on the graph. Provide the signs (positive or negative) of the partial derivatives of f at these points. For example, $\dfrac{\partial f}{\partial x} < 0$ at point A.

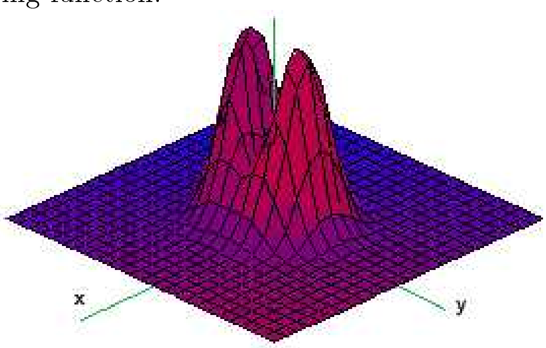


Exercise 4.12

Find all critical points of the function $f(x, y) = 2x^3 - 6x + y^2 - 2y + 7$.

Exercise 4.13

Make a sketch of contour (level) curves for the following function:



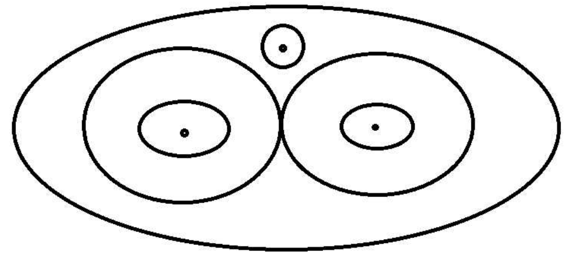
Exercise 4.14

The wave heights h in the open sea depend on the speed v of the wind and the length of time t that the wind has been blowing at that speed. Values of the function $h = f(v, t)$ are recorded in the table below. Estimate the rate of change of h with respect to v when $v = 40$ and $t = 15$. Show your computations.

$v\backslash t$	15	20	25
30	16	17	18
40	25	28	31
50	36	40	45

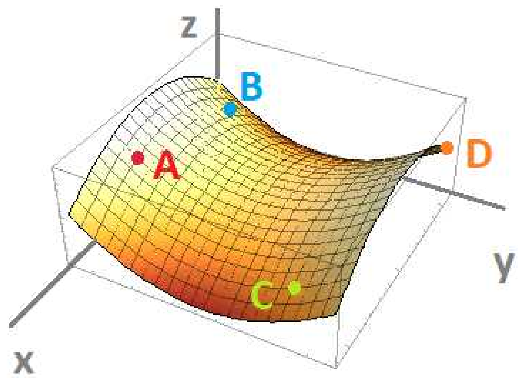
Exercise 4.15

The contour (level) curves for a function are given below. They are equally spaced. Sketch a possible graph that produced it and describe it.



Exercise 4.16

The graph of a function of two variables $z = f(x, y)$ is given below along with a point on the graph: 1. A, 2. B, 3. C, 4. D. Determine the signs of the derivatives f_x, f_{xx}, f_y, f_{yy} at that point:



Exercise 4.17

Show that the limit doesn't exist:

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2 + y^2}.$$

Exercise 4.18

Plot a few points of the graph of the function $h(x, y) = 3 - x - 2y$ to demonstrate that this is a plane.

Exercise 4.19

Find the best linear approximation of the function $f(x, y) = xy^2$ at the point $(0, 1)$.

Exercise 4.20

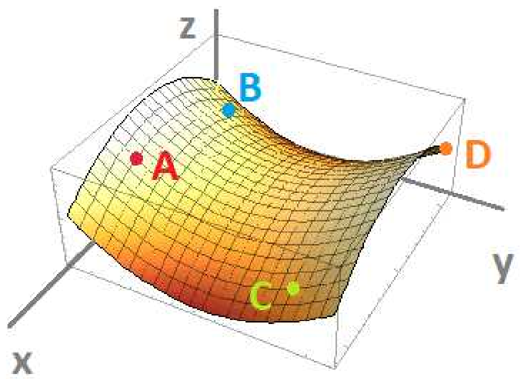
Find the linearization of the function $f(x, y) = e^{xy}$ at $(0, 1)$.

Exercise 4.21

Find the directional derivative of the function $f(x, y) = 2x^2 - 3y$ at the point $(1, 1)$ in the direction of the vector $< 1, 0 >$.

Exercise 4.22

Sketch the contour (level) curves of the function shown below, along with points A, B, C, D, on the xy-plane:



Exercise 4.23

Find the equation of the tangent plane to the surface $z = \sqrt{4 - x^2 - 2y^2}$ at the point $(1, -1, 1)$.

Exercise 4.24

Find all critical points of the function $f(x, y) = 2x^3 - 6x + y^2 - 2y + 7$.

Exercise 4.25

Find the directional derivative of the function $f(x, y) = x/(y + z)$ at the point $(4, 1, 1)$ in the direction of the vector $v =< 0, 2, -1 >$.

Exercise 4.26

Find the maximum rate of change of the function $f(x, y) = \sin(xy)$ at the point $(1, 0)$ and the direction in which it occurs.

Exercise 4.27

Find $\dfrac{dw}{dt}$, where

$$w = xy + yz^2\,,\ x = e^t\,,\ y = e^t \sin t\,,\ z = e^t \cos t\,.$$

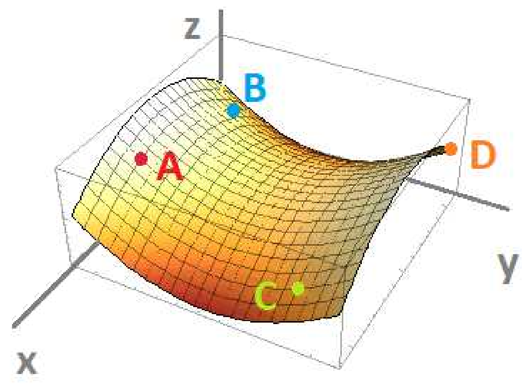
Exercise 4.28

Sketch the graph of a function of two variables $z = f(x, y)$ the derivatives of which have the following signs:

$$f_x > 0,\ f_{xx} > 0,\ f_y < 0,\ f_{yy} < 0\,.$$

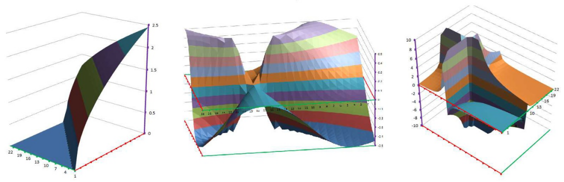
Exercise 4.29

The graph of a function of two variables $z = f(x, y)$ is given below along with four points on the graph. Sketch the gradient for each on a separate xy-plane:



Exercise 4.30

Make a sketch of contour (level) curves for the following function:



Exercise 4.31

Draw the contour map (level curves) of the following function of two variables:

1. $g(x, y) = \ln(x + y)$
2. $f(u, v) = uv$
3. $h(x, y) = 2x - 3y + 7$
4. $z = x^2 + y^2$

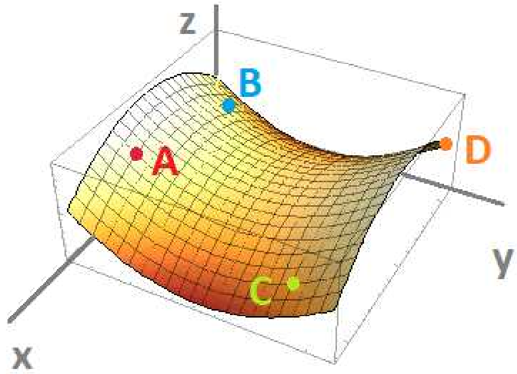
Exercise 4.32

Sketch the graph of a function of two variables $z = f(x, y)$ the derivatives of which have the following signs:

	1	2	3	4
f_x	$+$	$+$	$-$	$+$
f_{xx}	$-$	$+$	$+$	$-$
f_y	$-$	$-$	$+$	$+$
f_{yy}	$-$	$+$	$-$	$+$

Exercise 4.33

The graph of a function of two variables $z = f(x, y)$ is given below along with a point on the graph: 1. A, 2. B, 3. C, 4. D. Determine the signs of the derivatives f_x, f_{xx}, f_y, f_{yy} at that point:



Exercise 4.34

Find the partial derivatives of the following function of three variables:

1. $g(x, y, z) = z\sqrt{x + y}$
2. $u = 1$
3. $w = e^{x+y+z}$
4. $f(u, v, w) = uv + e^w$

Exercise 4.35

Find the best linear approximation of the function $f(x, y) = xe^y$ at the following point: 1. $(0, 0)$, 2. $(0, 1)$, 3. $(1, 1)$, 4. $(1, 0)$.

Exercise 4.36

From the definition, find the directional derivative of the function $f(x, y) = 2x - 3y$ at the point $(1, 1)$ in the direction of the following vector: 1. $< 3, 0 >$, 2. $-2\mathbf{j}$, 3. $< 1, 1 >$, 4. $< -1, -1 >$.

Exercise 4.37

Find the partial derivatives of the function

$$f(x, y) = \frac{x}{y} + \sin(xy) + xe^{2y}.$$

Exercise 4.38

Find the equation of the tangent plane to the surface $z = y\cos(x - y)$ at the point $(2, 2, 2)$.

Exercise 4.39

Set up as a max/min problem, but do not solve, the following: "Find the dimensions of a rectangular box of maximal volume such that the sum of lengths of its edges is equal to 10".

Exercise 4.40

State the chain rule, for the case of the composition of a function of two variables and a function of one variable.

Exercise 4.41

Find the linearization of the function $f(x, y) = e^{xy}$ at $(0, 1)$.

Exercise 4.42

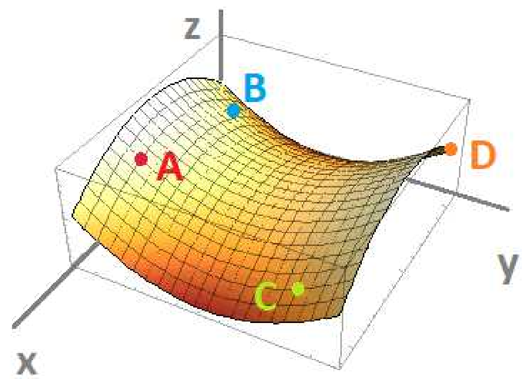
Find all critical points of the function $f(x, y) = 2x^3z - 6x + z^2 - 2y + 7$.

Exercise 4.43

Find the gradient of the function $f(x, y) = x^2y^{-3}$ at the point $(1, 1)$. Use this information to sketch the graph of f in the vicinity of this point. Explain.

Exercise 4.44

The graph of a function of two variables $z = f(x, y)$ is given below along with four points on the graph. Provide the signs of the partial derivatives of f at these points. For example, $\dfrac{\partial f}{\partial x} < 0$ at point A.



Exercise 4.45

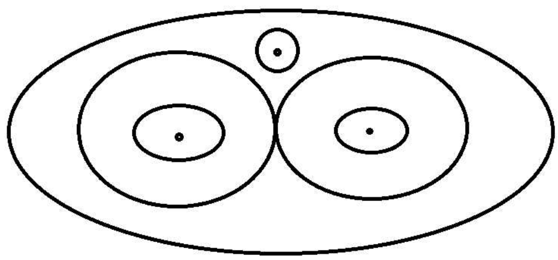
A function of two variables is given by $f(x, y) = xy^2$. (a) Find the linear approximation of f around the point $(1, 1)$. (b) Represent the graph of f as a parametric surface and find the plane tangent to this surface at the point $(1, 1, 1)$.

Exercise 4.46

State the chain rule, for the case of the composition of a function of two variables and a function of one variable.

Exercise 4.47

The contour (level) curves for a function are given below. They are equally spaced. Sketch a possible graph that produced it and describe it.

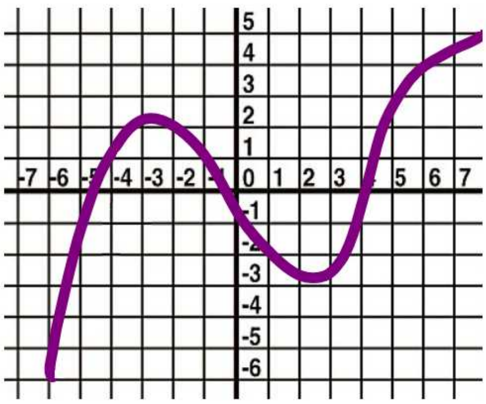


Exercise 4.48

Represent the function $h(x,y) = \sqrt{x^2y - 1}$ as the composition of two functions. Find its derivatives using the Chain Rule.

Exercise 4.49

The graph of function $y = g(x)$ of *one* variable is shown below. Suppose now that $z = f(x,y) = g(x)$ is a function of *two* variables, which depends only on x, given by the same formula. Find all points where the gradient of f is equal to 0.



Exercise 4.50

Find all critical points of the function $f(x,y) = 2x^3 - 6x + y^2 - 2y + 7$.

Exercise 4.51

Find the directional derivative of the function $f(x,y) = 2x^2 - 3y$ at the point $(1,1)$ in the direction of the vector $< 1, 0 >$.

Exercise 4.52

Represent the function $h(x,y) = \sqrt{x^2y - 1}$ as the composition of two functions. Find its derivatives using the Chain Rule.

Exercise 4.53

Find the partial derivatives of the following function of three variables:

1. $g(x,y,z) = z\sqrt{x+y}$
2. $u = 1$
3. $w = e^{x+y+z}$
4. $f(u,v,w) = uv + e^w$

Exercise 4.54

Find the best linear approximation of the function $f(x,y) = xe^y$ at the following point: 1. $(0,0)$, 2. $(0,1)$, 3. $(1,1)$, 4. $(1,0)$.

Exercise 4.55

From the definition, find the directional derivative of the function $f(x,y) = 2x - 3y$ at the point $(1,1)$ in the direction of the following vector: 1. $< 3, 0 >$, 2. $-2j$, 3. $< 1, 1 >$, 4. $< -1, -1 >$.

Exercise 4.56

A function of two variables is given by $f(x,y) = xy^2$. (a) Find the linear approximation of f around the point $(1,1)$. (b) Represent the graph of f as a parametric surface and find the plane tangent to this surface at the point $(1,1,1)$.

Exercise 4.57

Find the best linear approximation of the function $f(x,y) = xy^2$ at the point $(0,1)$.

Exercise 4.58

Find the maximum rate of change of the function $f(x,y) = \sin(xy)$ at the point $(1,0)$ and the direction in which it occurs.

Exercise 4.59

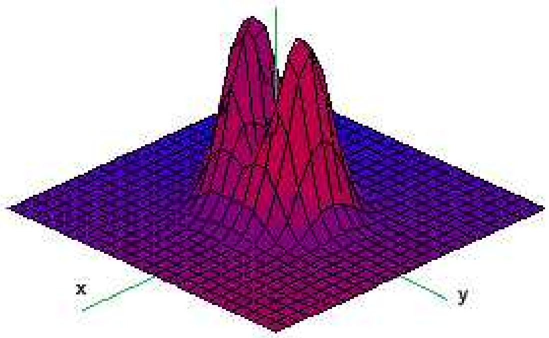
Find an equation of the tangent plane to the surface $z = 4x^2 - y^2 + 2y$ at the point $(-1,2,4)$.

Exercise 4.60

Find the directional derivative of the function $f(x,y) = 1 + 2x\sqrt{y}$ at the point $(3,4)$ in the direction of the vector $< 4, -1 >$.

Exercise 4.61

Make a sketch of contour (level) curves for the function below:



Exercise 4.62

Find the dimensions of a rectangular box of maximal volume such that the total surface area is equal to 64.

Exercise 4.63

A plane has an xyz-equation $2(x-1)+3(y-2)+4(z-3)=0$. Give a vector PARALLEL to the plane.

Exercise 4.64

Calculate the gradient $grad\, f(x,y,z)$ of $f(x,y,z)=xyz+\dfrac{x}{y}$.

Exercise 4.65

Does $f(x,y)=-x^2e^{x^2+y^2}$ have a maximmu or minimum? How do you know?

Exercise 4.66

Calculate $\dfrac{\partial f}{\partial z}$ for $f(x,y,z)=e^{xyz}+x$ at the point $(1,2,3)$.

Exercise 4.67

You are at the point $(0,0)$. In the direction of what vector should you step off $(0,0)$ in order to get the greatest initial increase in the function $f(x,y)=x^2+\dfrac{1}{4}y^2$. Explain.

Exercise 4.68

Let $f(x,y)=\sin(x-y)$. Give a formula for the function $h(x)$ defined by

$$h(x)=\int_0^x f(s,y)\,ds\,.$$

Exercise 4.69

Let

$$f(x,y)=\sqrt{x^2+(y-2)^2-4}+1\,.$$

1. Find the domain of f.
2. Sketch the graph of f.
3. What is it?

Exercise 4.70

Give an example of a function $z=f(x,y)$ such that $\dfrac{\partial f}{\partial x}(0,0)$ exists but $\dfrac{\partial f}{\partial y}(0,0)$ does not.

Exercise 4.71

From the definition, compute the directional derivative of the function

$$f(x,y)=xy+y^2$$

at the point $a=(2,1)$ in the direction $v=(1,1)$.

Exercise 4.72

From the definition, prove that $T(x)=-2x_2-1$ is the best affine approximation of the function

$$f(x)=x_1^4+x_2^2\,,$$

where $x=(x_1,x_2)$, at the point $a=(0,-1)$.

Exercise 4.73

Represent the function

$$f(t)=(\sin e^t,\cos e^t)$$

as the composition of two functions and then use the Chain Rule to compute its derivative at $t_0=0$.

Exercise 4.74

Find all critical points of the function

$$f(x,y,z)=xz+5y^2\,.$$

Exercise 4.75

Find and classify the critical points of the function $f(x,y)=x^4+y^4-4xy+1$.

Exercise 4.76

Find the tangent plane to the surface $f(x,y)=\sin(x-y)$ at $(\frac{\pi}{2},0,1)$.

Exercise 4.77

Give an example of a function of two variables such that:

1. Its image is the circle $x^2 + y^2 = 1$.
2. Its image is the sphere $x^2 + y^2 + z^2 = 1$.
3. Its graph is the sphere $x^2 + y^2 + z^2 = 1$.
4. Its domain is the disk $x^2 + y^2 \leq 1$.
5. It is equal to its own best linear (affine) approximation.

Exercise 4.78

Find the best affine approximation of the function

$$f(x) = x_1^2 + x_2^2 + x_3^2 \,,$$

where $x = (x_1, x_2, x_3)$, at the point $a = (0, -1, 1)$.

Exercise 4.79

Find the global minima of the function

$$f(x, y, z) = x^2 z^2 + y^2 + y \,.$$

Exercise 4.80

A mountain ridge has three peaks with two passes between them. Sketch the level curves of the function that represents the terrain.

Exercise 4.81

Define the gradient of a function. What does it tell us about the function?

5. Exercises: Integrals

Exercise 5.1

Estimate the integral

$$\iint_B x^2 y \, dA \,,$$

where

$$B = \{(x, y) : 0 \le x \le 2, 0 \le y \le 3\} \,,$$

by providing a Riemann sum with 6 squares.

Exercise 5.2

Use a Riemann sum with 8 terms to estimate the value of the integral

$$\iiint_D (x + y + z) \, dV \,,$$

over the cube $D = [0, 1] \times [0, 1] \times [0, 1]$. Choose your own sample points.

Exercise 5.3

Let $f : \mathbf{R} \to \mathbf{R}$ be a function such that $\int_a^b f(x) \, dx = 1$. Let $B = \{(x, y) : a \le x \le b, \ c \le y \le d\}$ be a rectangle. Find $\iint_B f \, dA$.

Exercise 5.4

Find the volume of the region bounded by the surface $z = 1 - x^2$, the xy-plane and the planes $y = 0$ and $y = 1$.

Exercise 5.5

Find the volume of the solid under the plane $x + 2y - z = 0$ and above the region bounded by $y = x$ and $y = x^4$.

Exercise 5.6

Find the area of the part of the surface $z = 1 + 3x + 2y^2$ that lies above the triangle with vertices $(0, 0)$, $(0, 1)$, and $(2, 1)$.

Exercise 5.7

Use a Riemann sum with $m = n = 2$ to estimate the value of the integral

$$\iint_D \sin(x + y) \, dA \,,$$

where $R = [0, \pi] \times [0, \pi]$. Choose your own sample points.

Exercise 5.8

Find the mass of the lamina that occupies the region D bounded by $y = e^x$, y=0, $x = 0$, $x = 1$, and its density function is $\rho(x, y) = y$.

Exercise 5.9

Express the length of the circle of radius r as an arc length integral and evaluate it.

Exercise 5.10

Evaluate the iterated integral

$$\int_0^1 \int_0^z \int_0^y z e^{-y^2} \, dx dy dz \,.$$

Exercise 5.11

Compute this integral below where region D is bounded by: $y = 0$, $y = x^2$, $x = 1$, $z = 0$, $z = 2$:

$$\iiint_D (y + z) \, dV \,.$$

Exercise 5.12

Find the volume of the solid under the plane $x + 2y - z = 0$ and above the region bounded by $y = x$ and $y = x^4$.

Exercise 5.13

Find the area of that part of the plane $z = 1 + 2x + 2y$ that lies directly above the region in the xy-plane bounded by the parabolas $y = x^2 - 1$ and $y = -x^2 + 1$.

Exercise 5.14

Find the area of the part of the surface $z = 1 + 3x + 2y^2$ that lies above the triangle with vertices $(0, 0)$, $(0, 1)$, and $(2, 1)$.

Exercise 5.15

Evaluate
$$\iiint_D 2x\, dV\,,$$
where
$$\begin{aligned} D \quad = \{(x,y,z):\ & 0 \le y \le 2\,,\\ & 0 \le x \le \sqrt{4-y^2}\,,\\ & 0 \le z \le y\}. \end{aligned}$$

Exercise 5.16

Evaluate the double integral
$$\iint_D (x+y)\, dA\,,$$
where D is bounded by $y=\sqrt{x}$ and $y=x^2$.

Exercise 5.17

Define the double integral of a function of two variables over a rectangle. Make a sketch and explain.

Exercise 5.18

Find the volume of the tetrahedron bounded by the planes $x+2y+z=2,\ x=2y,\ x=0$, and $z=0$.

Exercise 5.19

Define the double integral of a function of two variables over a rectangle. Make a sketch and explain.

Exercise 5.20

Find the volume of the tetrahedron bounded by the planes $x+2y+z=2,\ x=2y,\ x=0$, and $z=0$.

Exercise 5.21

Compute this integral
$$\iiint_D (y+z)\, dV\,,$$
where region D is bounded by: $y=0, y=x^2, x=1, z=0, z=2$.

Exercise 5.22

You are faced with a hand calculation of
$$\iint_R f(x)\, dxdy\,,$$
where R is the two-dimensional region consisting of everything bounded by the curves $y=x^2, y=x^2+1$ and the lines $x=0, x=1$. Switch to new, convenient for integration, coordinates u,v by indicating what u,v are in terms of x,y and describe R in terms of u,v.

Exercise 5.23

By hand calculation, evaluate
$$\iint_R 1\, dxdy\,,$$
where R is given in u,v coordinates as $0 \le u \le 1, 1 \le v \le 3$, and $x=u^2, y=v+u^2$.

Exercise 5.24

You are faced with a hand calculation of
$$\iiint_R f(x,y,z)\, dxdydz\,,$$
where R is the 3D region bounded from above by the unit sphere and from below by the xy-plane. Describe R in a way convenient for integration.

Exercise 5.25

You are faced with a hand calculation of
$$\iiint_R f(x,y,z)\, dxdydz\,,$$
where R is the 3D region consisting of everything bounded by the planes $y=x, y=x+1, y=-x+1, y=-x+2, z=0, z=2$. Switch to new, convenient for integration, coordinates u,v,w by indicating what u,v,w are in terms of x,y,z and describe R in terms of u,v,w.

Exercise 5.26

Find the volume conversion factor of the transformation
$$x(u,v,w)=u^2v\,,\ y(u,v,w)=v^2\,,\ z(u,v,w)=w^2e^u\,.$$

Exercise 5.27

You are faced with a hand calculation of
$$\iint_R f(x,y)\, dxdy\,,$$
where R is the two-dimensional region consisting of everything bounded by the curves $y=x^2, y=x^2+2$ and the lines $x+y=1, x+y=2$. (a) Sketch R in xy-plane. (b) Switch to new, convenient for integration, coordinates u,v by indicating what u,v are in terms of x,y. (c) Sketch R in uv-plane.

Exercise 5.28

By hand calculation, evaluate

$$\iint_R 2d\,xdy\,,$$

where R is given in u, v coordinates as $1 \leq u \leq 2, 0 \leq v \leq 1$, and $x = u^2 + 2v, y = ue^v$.

Exercise 5.29

You are faced with a hand calculation of

$$\iiint_R f(x, y, z)\,dxdydz\,,$$

where R is the "ice-cream cone", i.e., the 3D region obtained by intersecting the cone $z^2 = x^2 + y^2, z \geq 0$, and the sphere $x^2 + y^2 + z^2 = 2$. Describe R in a way convenient for integration.

Exercise 5.30

Find the volume conversion factor of the transformation

$$\begin{aligned} x(u, v, w) &= u^2 + v + 1 \\ y(u, v, w) &= v^2 + w + 2 \\ z(u, v, w) &= w^2 + u + 3 \end{aligned}$$

Exercise 5.31

Find by integration the area of that part of the plane $2x + 3y + z = 6$ that lies in the first octant.

Exercise 5.32

Find the volume of the solid in $\mathbf{R}^3$ bounded by the following surfaces: $z = x^2 + y^2$, $z = 0$, $y = x^2$, $y = 2x$.

Exercise 5.33

By means of an appropriate change of variables, evaluate the integral $\displaystyle\int_D xy\,dA$, where $D \subset \mathbf{R}^2$ is bounded by the lines $x = 0$, $x = 1$, $y = x$, $y = x+1$.

Exercise 5.34

Find the volume of a sphere of radius a.

Exercise 5.35

Find the volume of the region bounded by the surfaces $z = y$, $y = 4$, $z = 0$, $y = x^2$.

Exercise 5.36

Evaluate the integral

$$\iiint_B xyz\,dV\,,$$

where

$$B = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1\}\,.$$

Exercise 5.37

By using only the properties of the integral, compute

$$\iint_D f(x, y)\,dA\,,$$

where D is the disk $x^2 + y^2 \leq 4$ and $f(x, y) = 2$ if $-2 \leq x \leq 0$, $f(x, y) = -1$ if $0 \leq x \leq 2$, $f(x, y) = 55$ for all other values of (x, y).

Exercise 5.38

By using polar coordinates, compute the volume of the solid bounded by the cylinder $x^2 + y^2 = 1$ and the planes $z = 0$ and $z = 1$.

Exercise 5.39

Compute the volume of the solid lying above the rectangle R in the xy-plane consisting of all points $(x, y, 0)$ with $-1 \leq x \leq 1, 0 \leq y \leq 1$ and bounded from above by the surface given by $z = x^2y$.

6. Exercises: Vector fields

Exercise 6.1

Sketch the vector field given below and estimate its line integral along the boundary of the square oriented counterclockwise (multiple answers are possible):

$$\begin{array}{ll} F(0,0) =< 1,1 > & F(.5,0) =< 0,1 > \\ F(1,0) =< 1,1 > & F(1,.5) =< -1,1 > \\ F(1,1) =< -1,0 > & F(.5,1) =< 0,0 > \\ F(0,1) =< 2,1 > & F(0,.5) =< -1,-1 > \end{array}$$

Exercise 6.2

In the formula of Green's Theorem shown below, identify all of its parts (such as "F is..."):

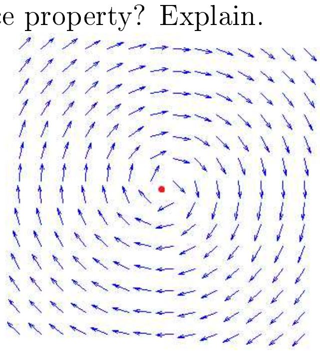
$$\oint_C F \cdot dP = \iint_D (q_x - p_y)\, dA\,.$$

Exercise 6.3

Sketch the velocity vector field $F(x,y) =< x, -y >$ identifying the most important features. Describe the motion in detail.

Exercise 6.4

State the path-independence property. Does the vector field shown below satisfy the path-independence property? Explain.



Exercise 6.5

(1) Represent the circle of radius 1 centered at 0 as a parametric curve. (2) Find the tangent line to this circle at the point $(\sqrt{2}/2, \sqrt{2}/2)$. (3) Compute the flux of the vector field $F =< 2,1 >$ across the part of the circle that lies in the first quadrant.

Exercise 6.6

Find the work done by force field

$$V(x,y) =< xy\,,\, y^2 >\,,$$

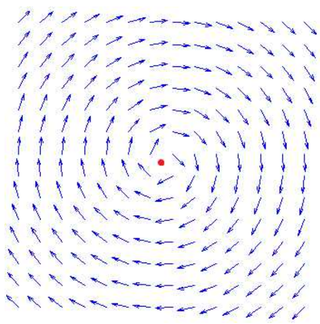
in moving an object along the parabola $x = t,\ y = t^2,\ 0 \le t \le 1$.

Exercise 6.7

Vector field V is sketched below. (a) Suppose C is the positively oriented curve following the upper half of the unit circle. Is

$$\int_C V \cdot dR$$

positive, negative or 0? Explain. (b) Suppose $P = (4,4)$. Is $\operatorname{div} V(P)$ positive, negative or 0? Explain.



Exercise 6.8

Evaluate the line integral

$$\int_C x^2 y^3\, dx - y\sqrt{x}\, dy\,,$$

where C is parametrized by $x = t^2,\ y = -t^3,\ 0 \le t \le 1$.

Exercise 6.9

(a) Explain what it means for a line integral $\int_C V \cdot dR$ to be independent of path. (b) Is $\int_C V \cdot dR$, where

$$V =< 1 - ye^{-x}\,,\, e^{-x} >$$

independent of path?

Exercise 6.10

(a) State Green's Theorem. (b) Verify it for the

vector field
$$V(x,y) = yi - xj$$
and a unit square.

Exercise 6.11

This is the formula of Green's Theorem:
$$\oint_C V \cdot dR = \iint_D \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} \right) dA\,.$$
Explain its parts and relations between them. Provide a sketch.

Exercise 6.12

Determine whether the vector field $V(x,y,z) = e^z i + j + xe^t k$ is conservative.

Exercise 6.13

Sketch the vector field
$$V(x,y) = \frac{1}{\sqrt{x^2+y^2}}(yi - xj)\,.$$

Exercise 6.14

Find the work done by force field
$$V(x,y) = xyi + y^2 j$$
in moving an object along the parabola $x = t,\ y = t^2,\ 0 \le t \le 1$.

Exercise 6.15

(a) Given a vector field $V(x,y) =< 3 + 2xy, x^2 - 3y^2 >$, find a function f such that $\nabla f = V$. (b) Use (a) to evaluate the line integral $\int_C V \cdot dR$, where C is the curve given by
$$R(t) =< e^t \sin t\,,\, e^t \cos t >,\ 0 < t < \pi\,.$$

Exercise 6.16

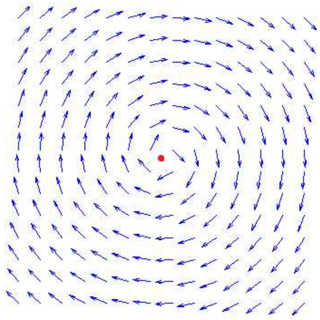
Verify Green's Theorem for the vector field
$$V(x,y) = (x-y)i + xj$$
and the region D bounded by the unit circle
$$C : R(t) = \cos ti + \sin tj,\ 0 \le t \le 2\pi\,.$$

Exercise 6.17

Sketch the vector field
$$\frac{1}{x^2+y^2} < x, y > \,.$$

Exercise 6.18

A vector field V is sketched below. Suppose C is the clockwise oriented square centered at the origin. Is $\int_C V \cdot dR$ positive, negative or 0? Explain.



Exercise 6.19

Prove that the vector field $V(x,y,z) = zj - yk$ is not conservative.

Exercise 6.20

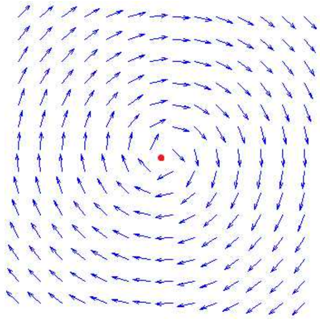
Sketch the vector field
$$V(x,y) = \frac{1}{x^2+y^2}(yi - xj)\,.$$

Exercise 6.21

(1) Represent the cylinder of radius 1 and height 1 centered on the z-axis as a parametric surface. (2) Find the tangent plane to the cylinder at the point $(\sqrt{2}/2, \sqrt{2}/2, 1/2)$. (3) Compute the flux of the vector field $F =< 2, 1, 1 >$ across the part of the cylinder that lies in the first octant.

Exercise 6.22

A vector field V is sketched below. Suppose C is the clockwise oriented square centered at the origin. Is $\int_C V \cdot dR$ positive, negative or 0? Explain.



Exercise 6.23

Prove that the vector field $V(x,y,z) = zj - yk$ is not conservative.

Exercise 6.24

Sketch the vector field

$$V(x,y) = \frac{1}{x^2+y^2}(yi - xj)\,.$$

Exercise 6.25

Find the work done by force field

$$V(x,y) =< xy, y^2 >$$

in moving an object along the parabola $x = t, y = t^2, 0 \le t \le 1$.

Exercise 6.26

Given vector field $F(x,y) = \left(x, \frac{1}{2}y\right)$. Choose a few points on the plane and draw the vector $F(x,y)$ with tail at (x,y). There are a few families of trajectories in this vector field. Pencil in a few trajectories of each type.

Exercise 6.27

Suppose that a mass M is fixed at the origin in space. When a particle of unit mass is placed at the point (x,y) other than the origin, it is subjected to a force $G(x,y)$ of gravitational attraction. Plot the vector field $G(x,y)$, if the magnitude (length) of $G(x,y)$ is $\frac{kM}{r^2}$, where $r = \sqrt{x^2+y^2}$.

Exercise 6.28

Here is a plot of a few trajectories (and vectors) of a vector field. On the basis of the plot, determine if it is a gradient field or not. Explain.

Exercise 6.29

Use Gauss's formula to evaluate the flow of the vector field $F(x,y,z) =< z^2, y, x^2 >$ across the surface of the pyramid bounded by the coordinate planes and the first octant part of the plane with equation $x+y+z=1$.

Exercise 6.30

Use Gauss's formula to evaluate the flow of the vector field $F(x,y,z) =< z+y, x+y, z >$ across the surface of the 3D box with a slanted top consisting of all points (x,y,z) with $0 \le x \le 1, 0 \le y \le 1, 0 \le z \le x+1$.

Exercise 6.31

Find the divergence and the rotation of the vector field $F(x,y) =< x^2y, xy\sin y >$.

Exercise 6.32

Measure the flow of the vector field $G(x,y) =< e^x + y, e^y >$ ALONG the boundary of the rectangle with corners at $(0,0),(1,0),(0,1),(1,1)$.

Exercise 6.33

Suppose a closed curve is located within the unit circle. Is the flow of $H(x,y) =< \frac{x^3}{3} - 2x, \frac{y^3}{3} >$ ACROSS this curve negative or positive? Explain.

Exercise 6.34

Find the best linear approximation of $F(x,y) = (xy, x^2+y^2)$ at the point where $x=1$ and $y=1$.

Exercise 6.35

Find the best linear approximation of the function $F(x,y,z) = (xz+y, x^2+zy^2)$ at the point where $x=1$, $y=1$, and $z=0$.

Exercise 6.36

Evaluate

$$\int_C (y^2+2xy)\,dx + (x^2+2xy)\,dy\,,$$

where C is the part of the graph $y = 2x^2$ from $(0,0)$ to $(1,2)$.

Exercise 6.37

Let $F(x,y) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$ be a vector field, and let C a simple (i.e., without self-intersections) closed path that encloses the origin. Find the work of F along C. Hint: It is equal to the work along a certain circle.

7. Examples

Example 7.1: function of two variables

Consider

$$f(x,y) = (x^2 + (y-2)^2 - y)^{\frac{1}{2}} + 1\,.$$

(a) Domain:

$$x^2 + (y-2)^2 - y \geq 0\,,$$
$$x^2 + (y-2)^2 \geq 4\,,$$

i.e., the region with boundary:

$$x^2 + (y-2)^2 = 2^2\,.$$

This is the circle of radius 2 with center at $(0,2)$.

(b) Graph:

$$f(0,y) = ((y-2)^2 - y)^{\frac{1}{2}} + 1 = z\,,$$

hence

$$z = (x^2 + (y-2)^2 - y)^{\frac{1}{2}} + 1 \implies (z-1)^2 = x^2 + (y-2)^2\,.$$

Then with $x = 0$:

$$(z-1)^2 = (y-2)^2 \implies z - 1 = \pm(y-2)\,,$$

and with $y = 0$:

$$(z-1)^2 = x^2 \implies z - 1 = \pm x\,.$$

Then $z = y - 1$ and $z = -y + 3$ result in two lines. These cross-sections give us the graph.
(c) This is funnel, a truncated cone.

Example 7.2: norm

To prove that $||x||$ is continuous, use

$$||x|| = (x_1^2 + ... + x_n^2)^{\frac{1}{2}}\,,$$

and show that

$$||x - a|| < \delta \implies \Big|||x|| - ||a||\Big| < \delta\,,$$

by the *Triangle Inequality*.

Example 7.3: union of path-connected

Suppose A and B are path-connected and a is in the intersection. If P and Q belong to the union, find a path from P to a, from a to Q. This gives you a path from P to Q.

Example 7.4: best linear approximation

Show that

$$T(x) = x_2^2 + 2x_2 + 1$$

is the best linear approximation of the function

$$f(x) = x_1^4 + x_2^2\,.$$

Consider

$$\begin{aligned}
\lim_{x\to a}\frac{f(x)-T(x)}{||x-a||} &= \lim_{x\to(0,-1)}\frac{x_1^4+x_2^2+2x_2+1}{||(x_1,x_2)-(0,-1)||}\\
&= \lim_{x\to(0,-1)}\frac{x_1^4+x_2^2+2x_2+1}{((x_1-0)^2+(x_2+1)^2)^{\frac{1}{2}}}\\
&= \lim_{x_2\to(-1)}\frac{x_2^2+2x_2+1}{|x_2+1|}\\
&= \lim_{x_2\to-1}\frac{(x_2+1)^2}{|x_2+1|} \qquad \text{by canceling } (x_2+1)\\
&= \lim_{x_2\to-1}|x_2+1|\\
&= 0.
\end{aligned}$$

Example 7.5: Chain Rule

Let

$$h(t)=\left(\sin e^t,\cos e^t\right).$$

Then h is the composition of

1. $f:\mathbf{R}\to\mathbf{R}$ given by $f(t)=e^t$, and
2. $g:\mathbf{R}\to\mathbf{R}^2, g(x)=(\sin x,\cos x)$.

Then

$$f'(t)=e^t, \nabla g=(\cos x,-\sin x),$$

and by the Chain Rule,

$$h'=(g\circ f)'=\nabla g\cdot f'=(\cos e^t,-\sin e^t)e^t.$$

Example 7.6: derivative of parametric curve

Describe the curve which results from the vector valued function

$$r(t)=(\cos 2t,\sin 2t,t),$$

where $t\in\mathbf{R}$.
Solution: The first two components indicate that for

$$r(t)=(x(t),y(t),z(t)),$$

the pair $(x(t),y(t))$ traces out a circle. While it is doing so, $z(t)$ is moving at a steady rate in the positive direction. Therefore, the curve which results is a cork screw shape, i.e. a helix.

Example 7.7: particle

The position of a particle at time t is (x,y), where

$$x=\sin\ t,$$

$$y=\sin^2 t.$$

Describe the motion of the particle as t varies over the time interval $[a,b]$.
Solution: We can eliminate t to see that the motion of the object takes place on the parabola $y=x^2$. The orientation of the curve is from $(\sin a,\sin^2 a)$ to $(\sin b,\sin^2 b)$.

Example 7.8: limit

Find the limit

$$\lim_{x\to 1}\frac{x^2-x}{x-1}.$$

Solution: It is

$$\frac{x \cdot (x-1)}{x-1} = x \text{ for } x \neq 1\,.$$

Hence

$$\lim_{x \to 1} \frac{x \cdot (x-1)}{x-1} = \lim_{x \to 1} x = 1\,.$$

Example 7.9: limit at infinity

Find the limit

$$\lim_{x \to \infty} \frac{x}{1+x}\,.$$

Solution: Rewrite

$$\frac{x}{1+x} = \frac{1}{1+\frac{1}{x}}\,.$$

Now it is

$$\lim_{x \to \infty} \frac{1}{1+x} = 1 \neq 0\,.$$

Therefore,

$$\lim_{x \to \infty} \frac{x}{1+x} = \lim_{x \to \infty} \frac{1}{1+\frac{1}{x}} = \frac{1}{1} = 1\,.$$

Example 7.10: continuity on plane

Let $f : \mathbf{R} \to \mathbf{R}^2$. Is

$$f(t) = ((t+1)^{\frac{1}{2}}, \tan(t))$$

continuous?
Solution: The first component is defined and continuous for

$$t \geq -1,$$

while the second component is defined and continuous for

$$t \neq \frac{\pi}{2} + k\pi, k \in \mathbf{Z}\,.$$

Hence it is continuous on $\mathbf{R}$ for

$$t \geq -1 \text{ and } t \neq \frac{\pi}{2} + k\pi, k \in \mathbf{Z}\,.$$

Example 7.11: continuity in space

Let $f : \mathbf{R} \to \mathbf{R}^3$,

$$f(t) = \left(\frac{1}{t^2-1}, (1-t^2)^{\frac{1}{2}}, \frac{1}{t}\right)\,.$$

If f continuous?

Solution: The first component is defined and continuous for

$$t \neq \pm 1\,.$$

The second component is defined and continuous for

$$t \in [-1, 1]\,,$$

while the third component is defined and continuous for

$$t \neq 0\,.$$

Hence, f is continuous on $\mathbf{R}$ for

$$t \in (-1, 0) \cup (0, 1)\,.$$

Example 7.12: dimension 4

Let $g : \mathbf{R} \to \mathbf{R}^4$,

$$g(t) = (\cos(4t), 1 - (3t+1)^{\frac{1}{2}}, \sin(5t), \sec(t)) .$$

If g continuous?

Solution: The first and the third component are continuous on $\mathbf{R}$. The second component is continuous and defined on $(-\frac{1}{3}, \infty)$, while the secant $\sec(t) = \dfrac{1}{\cos(t)}$ in the fourth component is defined and continuous for

$$t \neq \frac{\pi}{2} + k\pi, k \in \mathbf{Z} .$$

Hence, $g(t)$ is continuous as long as

$$t \geq -\frac{1}{3} \text{ and } t \neq \frac{\pi}{2} + l\pi,\ l \in \mathbf{Z} .$$

Example 7.13: tangent vector

Let

$$f(t) = (\sin\ t, t^2, t+1) \text{ for } t \in [0,5] .$$

Find a tangent line to the curve parametrized by f at the point $t = 2$.

Solution: A direction vector has the same direction as $f'(2)$. Therefore, it suffices to simply use $f'(2)$ as a direction vector for the line. Further,

$$f'(2) = (\cos\ 2, 4, 1) .$$

Hence, a parametrized equation for the tangent line is

$$(x, y, z) = (\sin\ 2, 4, 3) + t(\cos\ 2, 4, 1) .$$

Example 7.14: velocity vector

Let

$$f(t) = (\sin t, t^2, t+1) \text{ for } t \in [0,5] .$$

Find the velocity vector for $t = 1$.

Solution: The velocity vector is simply $f'(1) = (\cos 1, 2, 1)$.

Example 7.15: velocity-speed-acceleration

Consider $g : \mathbf{R} \to \mathbf{R}^2$,

$$g(t) = (t, t^2) .$$

What is its velocity, speed, and acceleration as it passes through $(2,4)$?

Solution: It is

$$g'(t) = (1, 2t)$$

and

$$g'`(t) = (0, 2) .$$

Further, at $(2,4)$, we know $t = 2$. For the velocity at $t = 2$, we have $g'(2) = (1,4)$. The speed equals

$$||g'(2)|| = (1^2 + 4^2)^{\frac{1}{2}} = \sqrt{17} ,$$

whereas the acceleration is equal to $g'`(2) = (0, 2)$.

Example 7.16: velocity

An object has position

$$r(t) = (t^3, \frac{t}{1+t}, (t^2+2)^{\frac{1}{2}})\ km\,.$$

where t is given in hours. Find the velocity of the object in kilometers per hour when $t = 1$.

Solution: Since velocity at time t equals $r'(t)$, we calculate

$$\begin{aligned} r'(t) &= \left(3t^2, \frac{1(1+t)-t}{(1+t)^2}, 2t \cdot \frac{1}{2}(t^2+2)^{-\frac{1}{2}}\right) \\ &= \left(3t^2, \frac{1}{(1+t)^2}, \frac{1}{(t^2+2)^{\frac{1}{2}}}t\right). \end{aligned}$$

For $t = 1$, the velocity is

$$r'(1) = \left(3, \frac{1}{4}, \frac{1}{\sqrt{3}}\right) km/hour\,.$$

Example 7.17: hyperbola

Consider $g : \mathbf{R} \to \mathbf{R}^2$,

$$g(t) = (\frac{1}{t}, 2t)$$

defined on $(0, \infty)$. Its image is one branch of a hyperbola (this can be seen by writing $g(t) = (x, y)$, i.e. $x = \frac{1}{t}, y = 2t$, which yields $y = \frac{2}{x}$). Find the velocity, the speed and the acceleration at time t.

Solution. The velocity equals

$$g'(t) = (-\frac{1}{t^2}, 2)\,,$$

the acceleration is equal to

$$g'`(t) = (\frac{2}{t^3}, 0)\,,$$

and the speed is

$$||g'(t)|| = ||(-\frac{1}{t^2}, 2)|| = (\frac{1}{t^4} + 4)^{\frac{1}{2}}\,.$$

Example 7.18: conjecture

For the function

$$f(t) = (3t\cos(2t), 4t\sin(2t))\,,$$

show that

$$\frac{d}{dt}||f(t)|| \neq ||f'(t)||\,.$$

Solution: It is

$$||f(t)|| \quad = (9t^2\cos^2(2t) + 16t^2\sin^2(2t))^{\frac{1}{2}} \quad = (9t^2 + 7t^2\sin\}^2(2t))^{\frac{1}{2}},$$

and its derivative equals

$$\frac{d}{dt}||f(t)|| = \frac{9t + 14t^2\sin(2t)\cos(2t) + 7t\sin^2(2t)}{9t^2 + 7t^2\sin^2(2t))^{\frac{1}{2}}}\,.$$

Further

$$f'(t) = (3\cos(2t) - 6t\sin(2t), 4\sin(2t) + 8t\cos(2t))\,,$$

and thus

$$\begin{aligned} ||f'(t)|| &= ((3\cos(2t) - 6t\sin(2t))^2 + (4\sin(2t) + 8t\cos(2t))^2)^{\frac{1}{2}} \\ &= (9 + 7\sin^2(2t) + 36t^2 + 28t^2\cos^2(2t) + 14t\sin(4t))^{\frac{1}{2}}. \end{aligned}$$

Now evaluating both functions at $t = \pi$ (for example) yields

$$(\frac{d}{dt}||f(t)||)_{t=\pi} = \frac{9\pi}{\sqrt{9\pi^2}} = 3\,,$$

whereas

$$||f'(\pi)|| = \sqrt{9 + 64\pi^2} \approx 25.31\,,$$

hence $\frac{d}{dt}||f(t)|| \neq ||f'(t)||$.

Example 7.19: flow line

Let

$$F = yi - xj + 2k =< y, -x, 2 >$$

be a vector field. Verify that the path

$$g(t) = (\sin t, \cos t, 2t)$$

is a flow line (i.e. a path such that the velocity along the path is a vector in the vector field, $g'(t) = F(g(t))$) for the vector field F.

Solution: With

$$(x, y, z) = (\sin t, \cos t, 2t)$$

it follows that

$$g'(t) = (\cos t, -\sin t, 2)\,,$$

and

$$F(g(t)) = (y, -x, 2t) = (\cos t, -\sin t, 2t)\,.$$

Example 7.20: astroid

Consider the path $g : \mathbf{R} \to \mathbf{R}^2$,

$$g(t) = (\cos^3 t, \sin^3 t)\,,$$

which describes an astroid. For $t \in [0, \frac{\pi}{2}]$, one fourth of the astroid is described. Calculate its arc-length.

Solution: With

$$g'(t) = (-3 \sin t \cos^2 t, 3 \sin^2 t \cos t)\,,$$

$$||g'(t)|| = 3 \sin t \cos t = \frac{3}{2} sin2t\,,$$

it is

$$\begin{aligned} \int_0^{\frac{\pi}{2}} ||g'(t)||\, dt \qquad &= \int_0^{\frac{\pi}{2}} \frac{3}{2} \sin 2t\, dt \\ &= -\frac{3}{4} \cos\}(2t)|_0^{\frac{\pi}{2}} \\ &= -\frac{3}{4} \cos \pi + \frac{3}{4} \cos 0 = \frac{3}{2}. \end{aligned}$$

Example 7.21: arc-length

Consider the path $f : \mathbf{R} \to \mathbf{R}^2$,

$$f(t) = (t^2, \frac{2}{3}(2t+1)^{\frac{3}{2}}), 0 \leq t \leq 4\,.$$

Calculate its arc-length.

Solution: It is

$$\int_0^4 ||f'(t)||dt \quad = \int_0^4 (4t^2 + 4(2t+1))^{\frac{1}{2}} dt \quad = \int_0^4 2(t+1)dt \quad = t^2 + 2t|_0^4 \quad = 24.$$

Example 7.22: helix

Find the arc-length of the helix

$$g(t) = (\cos t, \sin t, t)$$

from $t = 0$ to $t = 2\pi$.

Solution: The arc-length equals

$$\begin{aligned}\int_0^{2\pi} ||g'(t)||dt &= \int_0^{2\pi} (-(\sin t)^2 + \cos^2 t + 1)^{\frac{1}{2}}\, dt \\ &= \int_0^{2\pi} (\sin^2 t + \cos^2 t + 1)^{\frac{1}{2}} dt \\ &= \int_0^{2\pi} \sqrt{2} dt \\ &= \sqrt{2}(2\pi - 0) \\ &= 2\sqrt{2}\pi.\end{aligned}$$

Example 7.23: arc-length parametrization

Parametrize the helix

$$g(t) = (\cos t, \sin t, t)$$

for $t \in [0, 2\pi]$ by arc-length.

Solution: It is

$$\int_0^t ||g'(t)||dt = \int_0^t \sqrt{2}dt = \sqrt{2}t$$

for all $t \in [0, 2\pi]$. We can solve for t in terms of s:

$$t = \alpha(s) = \frac{s}{\sqrt{2}}\,,$$

and hence

$$g(s) = \left(\cos \frac{s}{\sqrt{2}}, \sin \frac{s}{\sqrt{2}}, \frac{s}{\sqrt{2}}\right)$$

for all $s \in [0, 2\sqrt{2}\pi]$.

Example 7.24: cylindrical coordinates

Find the arc-length of the curve whose cylindrical coordinates are given by

$$r = e^t\,,$$

$$\theta = t\,,$$

$$z = e^t$$

for $t \in [0, 1]$.

Solution: With

$$r'(t) = e^t,\ \theta'(t) = 1,\ z'(t) = e^t\,,$$

we obtain

$$\begin{aligned}\int_0^1 (r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2)^{\frac{1}{2}}\, dt &= \int_0^1 (e^{2t} + e^{2t}(1) + e^{2t})^{\frac{1}{2}}\, dt \\ &= \int_0^1 e^t\sqrt{3}\, dt \\ &= \sqrt{3}(e-1).\end{aligned}$$

Example 7.25: gradient

Interpretation of the gradient. Consider a room in which the temperature is given by a scalar field T, so that at each point (x, y, z) the temperature equals $T(x, y, z)$, assuming the temperature does not change in time. How can the gradient be interpreted?

Solution: In this case, at each point in the room, the gradient of T at that point will show the direction in which the temperature rises most quickly. The magnitude of the gradient will determine how fast the temperature rises in that direction.

Example 7.26: gradient

Interpretation of the gradient. Consider a hill whose height above sea level at a point (x, y) is $H(x, y)$. How can the gradient of H be interpreted?

Solution: The gradient of H at a point is a vector pointing in the direction of the steepest slope at that point. The steepness of the slope at that point is given by the magnitude of the gradient vector.

Example 7.27: fly

Consider a room in which the temperature is given by

$$f(x, y, z) = xy^2z^3\,.$$

Consider further that a fly is crawling at unit speed in the direction of the vector

$$v =< -1, 1, 0 >$$

starting at the point

$$s = (2, 1, 1)\,.$$

Compute that rate of temperature change the fly is about to experience.

Solution: It is

$$\nabla f(x, y, z) =< y^2z^3, 2xyz^3, 3xy^2z^2 >\,,$$

hence

$$\nabla f(2, 1, 1) =< 1, 4, 6 > \ .$$

Further,

$$\frac{v}{||v||} =< -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 >\,,$$

and thus

$$\frac{df}{ds} =< 1, 4, 6 > \cdot < -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 >= \frac{3}{\sqrt{2}}\,.$$

Example 7.28: directional derivative

Let $f : \mathbf{R}^3 \to \mathbf{R}$,

$$f(x, y, z) = x + \sin(xy) + z\,.$$

Find the directional derivative $D_v f(1, 0, 1)$, where

$$v =< \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} > \ .$$

Solution: Note that v is already a unit vector. Therefore, it is only necessary to find $\nabla f(1, 0, 1)$ and take the dot product. It is

$$\nabla f(x, y, z) = (2x + \cos(xy)y, \cos(xy)x, 1)\,,$$

and hence

$$\nabla f(1, 0, 1) = (2, 1, 1)\,.$$

The directional derivative is obtained as

$$< 2,1,1 > \cdot < \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} >= \frac{4}{3}\sqrt{3}\,.$$

Example 7.29: tangent plane

Let $f : \mathbf{R}^3 \to \mathbf{R}$. Find the equation of the tangent plane to the level surface $f(x, y, z) = 6$ of the function

$$f(x, y, z) = x^2 + 2y^2 + 3z^2$$

at the point $(1, 1, 1)$.

Solution: First note that $(1, 1, 1)$ is a point on the level surface (i.e. $f(1, 1, 1) = 6$). To find the desired plane it suffices to find the normal vector to the proposed plane. But we see

$$\nabla f(x, y, z) =< 2x, 4y, 6z >\,,$$

and hence

$$\nabla f(1, 1, 1) =< 2, 4, 6 > \ .$$

Therefore, the equation of the tangent plane is:

$$< 2, 4, 6 > +(x - 1, y - 1, z - 1) = 0\,,$$

or

$$2x + 4y + 6z - 12 = 0\,.$$

Example 7.30: gradient

Compute the gradient of the function

$$f(x, y) = x^2 \sin(xy)$$

at $(\pi, 0)$.

Solution: It is

$$\nabla f(\pi, 0) = \begin{bmatrix} f_x(\pi, 0) \\ f_y(\pi, 0) \end{bmatrix} = \begin{bmatrix} 2x\sin(xy) + x^2 y \cos(xy) \\ x^3 \cos(xy) \end{bmatrix}_{(\pi,0)} =< 0, \pi^3 > \ .$$

Next...

CALCULUS ILLUSTRATED

VOLUME 5: DIFFERENTIAL EQUATIONS

PETER SAVELIEV

Index